

2015

Investigation into the Specification of NURF Recruitment to the Genome

Marissa Mack

Virginia Commonwealth University, marissaimack@gmail.com

Follow this and additional works at: <http://scholarscompass.vcu.edu/etd>

 Part of the [Genetics Commons](#)

© The Author

Downloaded from

<http://scholarscompass.vcu.edu/etd/3843>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

INVESTIGATION INTO THE SPECIFICATION OF NURF RECRUITMENT TO THE GENOME

A thesis submitted in partial fulfillment of the requirement for the degree Master of
Science at Virginia Commonwealth University

by

Marissa Irene Mack
Bachelor of Science, Binghamton University, 2011

Director: Joseph W. Landry, Ph.D.
Assistant Professor
Department of Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
April 2015

Acknowledgement

I would like to thank my advisor, Dr. Joseph Landry, for sharing his passion for science with me. He has pushed me to become a better scientist, and he is someone that I will continue to respect and admire for their hard work and dedication to the field of science.

I would also like to thank my committee members, Dr. Shirley Taylor and Dr. Joyce Lloyd. They provided me with great advice and support through this project.

This project would not be what it is without the help of Charles Lyons, a member of Dr. Richard Moran's lab. He was an important resource for the Mass Spectrometry experiments, and was incredibly patient with helping me understand the technique.

The members of the Landry lab were also crucial to the completion of this project. Aiman Alhazmi, Kimberly Mayes, and Kristen Peterson were invaluable to me in their support through this project. I would like to specifically thank Aiman, for his endless assistance and working with me for the past two years.

I would like to thank my family for their endless support. My parents continually inspire me with their hard work. They both have taught me that hard work and dedication can take you anywhere you want to go. I also need to thank my fiancé for his incredible support. He kept me grounded throughout this entire process, and reminded me to step back and breath.

Table of Contents

	Page
Acknowledgement	ii
List of Tables and Figures	vi
List of Abbreviations	viii
Abstract	xiii
Chapter 1: Introduction	1
Epigenetics.....	1
Chromatin Remodeling.....	5
Nucleosome Remodeling Factor.....	8
BPTF.....	10
PHD Finger Domains.....	13
Rationale for Study.....	16
Specific Aims.....	17
Chapter 2: Materials and Methods	18
Cell Culture.....	18
Cloning.....	19
Recombinant Protein Expression and Purification.....	21
Pull Downs.....	22

1. Nucleic Acid Pull Downs.....	22
2. Histone Pull Down.....	23
3. Peptide Pull Down.....	24
Affinity Column.....	25
1. Protein Expression.....	25
2. Protein Purification.....	25
3. Protein Coupling.....	26
4. Nuclear Extract.....	27
5. Pull Down.....	27
6. Mass Spectrometry.....	29
Nucleofection.....	30
Protein Extraction.....	31
Protein Quantification.....	31
mRNA Purification.....	32
Chapter 3: Results.....	34
Addition of two Exons Does not Enhance the Exogenous Expression of Bptf cDNA In Mouse ES Cells.....	34
The N-Terminal PHD Finger Domain of Bptf Does Not Interact With Histones.....	37
The N-Terminal PHD Finger Domain of Bptf Does not Interact With Nucleic Acids.....	40
The N-Terminal PHD Finger Domain of Bptf is Unlike Other PHD Finger	

Domains.....	43
The N-Terminal PHD Finger Domain of Bptf Interacts With Nonhistone Proteins.....	46
Chapter 4: Discussion.....	53
The Bptf cDNA is Unable to be Expressed in Mouse ES Cells.....	53
The N-Terminal PHD Finger Domain of Bptf is not a Standard PHD Finger Domain.....	55
Two Potential Binding Partners for Bptf Identified.....	59
Conclusions.....	62
Literature Cited.....	65
Vita.....	72

List of Tables and Figures

	Page
Figure 1. Cartoon Representation of Chromatin Compaction in the Nucleus of a Cell...	3
Figure 2. Electron Microscopy of Chromatin Fiber.....	4
Figure 3. Cartoon Representation of the Functions of Chromatin Remodeling Complexes.....	6
Figure 4. Cartoon Representation of the <i>Drosophila</i> and Mammalian NURF Complexes.....	9
Figure 5. Functional Domains of BPTF.....	12
Figure 6. Spatial Representation of the PHD Finger Domain in Pygopus.....	15
Figure 7. Schematic of Cloning Strategy for Bptf cDNA.....	20
Figure 8. Schematic of the Two Affinity Columns Generated.....	26
Figure 9. Visual of GFP Expression in CJ7 Cells.....	35
Figure 10. The Bptf cDNA is not Expressed in ES Cells.....	36
Figure 11. N-PHD Does not Recognize Select Histone Modifications.....	38
Figure 12. Bptf's N-PHD Domain Does not Interact With Histones.....	39
Figure 13. The N-Terminal PHD Finger Domain of Bptf Does not Interact With DNA...	41
Figure 14. The N-Terminal PHD Finger Domain of Bptf Does not Interact With RNA...	42
Figure 15. Alignment of the N-Terminal PHD Finger Domain Across 18 Species.....	44

Figure 16. The N-PHD is not Similar to Other PHD Finger Domains.....	45
Figure 17. The N-PHD Domain Interacts With Nuclear Proteins.....	48
Figure 18. Representative MS ² Peptide Scan for Thoc4.....	50
Figure 19. Representative MS ² Peptide Scan for Hmgb1.....	51
Figure 20. Hmgb1 and Thoc4 are Specific to N-PHD.....	52
Figure 21. Models for Interaction Between NURF and Thoc4 (A) and Hmgb1 (B).....	62
Table 1. Primers Used in Generation of GST/MBP Fusion Proteins, as Well as the Amplification of New Sequence in Bptf.....	33
Table 2. Proteins Identified as Potential Interaction Partners.....	49

List of Abbreviations

AB: Ammonium Bicarbonate

Ap-1: Activator Protein 1

AT: Adenine-Thymine

ATP: Adenosine Triphosphate

BCL9: B-Cell CLL/Lymphoma 9 Protein

BP: Base Pairs

BPTF: Human Bromodomain PHD Finger Transcription Factor

Bptf: Mouse Bromodomain PHD Finger Transcription Factor

BRPF2: Bromodomain and PHD Finger Containing Protein 1

C-PHD: C-terminal PHD finger of Bptf

cDNA: Complementary DNA

CHD: Chromodomain Helicase Binding Protein

ChIP: Chromatin Immunoprecipitation

cm: Centimeter

CpG: Cytosine Guanine

CRISPR: Clustered Regularly Interspersed Short Palindromic Repeats

CTCF: CCCTC Binding Factor

DDT: DNA binding homeobox and Different Transcription factors

DMEM: Dulbecco's Modified Eagle Medium

DNA: Deoxyribonucleic Acid

DNAse I: Deoxyribonuclease I

DNMT: DNA Methyl Transferase

Dpp: Decapentaplegic

DTT: Dithiothreitol

DVE: Distal Visceral Endoderm

EDTA: Ethylenediaminetetraacetic acid

EMT: Epithelial Mesenchymal Transition

ERK: Extracellular Signal Regulated Kinase

ESC: Embryonic Stem Cells

FBS: Fetal Bovine Serum

G: Gravity

GFP: Green Fluorescent Protein

GST: Glutathione S Transferase

H1: Histone protein 1

H2A: Histone protein 2A

H2B: Histone protein 2B

H3: Histone protein 3

H3K4me3: Histone 3 Lysine 4 Trimethyl

H3K9me3: Histone Protein 3 Lysine 9 Trimethyl

H4: Histone Protein 4

H4K16Ac: Histone 4 Lysine 16 Acetyl

HMGA: High Mobility Group Protein A

Hmgb1: High Mobility Group Protein B1

HRP: Horseradish Peroxidase

ING2: Inhibitor of Growth Protein 2

INO80: Inositol Requiring Mutant 80

ISWI: Imitation Switch

JAK/STAT: Janus Kinase/Signal Transducer and Activator of Transcription

K12: Lysine 12

K16: Lysine 16

K20: Lysine 20

kDa: Kilo Dalton

L: Liter

LIF: Leukemia Inhibitory Factor

M: Molar

mA: Milliamps

MAPK: Mitogen-Activated Protein Kinase

MBP: Maltose Binding Protein

mL: Milli Liter

mM: Milli Molar

MNase: Micrococcal Nuclease

mRNA: Messenger RNA

N-PHD: N-terminal PHD finger of Bptf

NFDM: Non-fat Dry Milk

ng: Nanogram

nm: Nanometer

NURF: nucleosome remodeling factor

OD: Optical Density

p300: Protein 300

PBS: Phosphate Buffered Saline

PBST: Phosphate Buffered Saline Tween-20

PCR: Polymerase Chain Reaction

PDB: Pull Down Buffer

PHD: Plant Homeodomain

PR: Progesterone Receptor

PVDF: Polyvinylidene Difluoride

Q: Glutamic Acid

RNA: Ribonucleic Acid

RPM: Revolutions Per Minute

SDS: Sodium Dodecyl Sulfate

Smad: Mothers Against Decapentaplegic, Drosophila homolog

SNF2L: Sucrose NonFermentable Protein 2L

SRF: Serum Response Factor

SRSF9: Serine Rich Splicing Factor 9

SWI/SNF: Switch/Sucrose Non-Fermentable

TGF β : Transforming Growth Factor Beta

Thoc4: THO Complex Subunit 4

TOPO: DNA Topoisomerase I

TREX: Transcription Export

Tris: Tris(hydroxymethyl)aminomethane

V: Volts

VCU: Virginia Commonwealth University

VE: Visceral Endoderm

X: Times

Zn: Zinc

γ -H2AX: gamma Histone protein 2 A.X

μ g: Microgram

μ L: Microliter

Abstract

INVESTIGATION INTO THE SPECIFICATION OF NURF RECRUITMENT TO THE GENOME

by Marissa Irene Mack, B.S

A thesis submitted in partial fulfillment of the requirement for the degree Master of Science at Virginia Commonwealth University

Virginia Commonwealth University, 2015

Major Director: Dr. Joseph W. Landry
Assistant Professor
Department of Human and Molecular Genetics

The nucleosome remodeling factor (NURF) is a multi-protein complex that plays a role in the regulation of gene expression through its ability to remodel nucleosomes. The largest subunit of this complex, Bptf (Bromodomain PHD Finger Transcription Factor) is important for many cellular processes as a transcriptional regulator and improper function results in disease or malignancy. To further understand the genome-wide recruitment of the NURF complex, the interaction partner for the N-terminal PHD finger domain of Bptf was investigated through pull down assays followed by mass spectrometry. It was determined that this domain does not recognize histones; instead it

recognizes a nonhistone protein, Thoc4 or Hmgb1. The expression of a cDNA corresponding to Bptf was also tested for expression in mouse ES cells after the addition of two exons found to be missing in the original cDNA. Addition of this sequence did not allow for exogenous Bptf expression in ES cells.

Chapter 1: Introduction

Epigenetics

Epigenetics is a rapidly expanding field in biology that can be defined as a heritable phenotype that is not the result of changes in the DNA sequence (27). Through the process of development, a single cell divides to give rise to many different cell types, and this is achieved in part through epigenetic mechanisms (9). The state of chromatin in the cell plays a role in cellular processes that involve DNA such as DNA replication, repair, and transcription (9). The state of chromatin is mutable and controls the access of transcription factors and other DNA associated proteins to the DNA at different moments of cellular metabolism (4). More open chromatin, euchromatin, provides better access to DNA for specific proteins and ultimately promotes gene expression, while condensed chromatin, heterochromatin, restricts access to the DNA and promotes gene silencing (18). In part, the expression of genes is determined by the chromatin structure of a cell (9).

Chromatin is the cell's solution to the large amount of DNA found in each cell (Figure 1). The total length of DNA in a human cell is around two meters long if it is stretched out, however the nucleus is very small with a diameter around 6 μm (4). Chromatin structure is composed of a nucleoprotein complex that involves ~147 bp of DNA wrapped around a histone protein octamer consisting of two copies of each core

histone protein H2A, H2B, H3 and H4 (9). This nucleoprotein complex is called the nucleosome. There is a linker sequence of DNA between nucleosomes that contains between 10 – 80 nucleotides that connects nucleosomes and this string of DNA and nucleosomes represents the “beads on a string” model as seen in Figure 2B (4). From here, the DNA is further condensed into a 30nm fiber, and further into the familiar chromosome.

One documented epigenetic mechanism involves histone modifications. The histone proteins that comprise the histone protein octamer are small, highly conserved proteins that are highly abundant in cells (4). The size of histone proteins ranges from 102 to 135 amino acids and they have a shared motif called the histone fold (4). This histone fold allows histones to form dimers and come together to generate the histone octamer (4). The histone proteins are lysine and arginine rich, generating a positive charge to help neutralize the negative charge of the DNA backbone. The N-terminal tail of histones extends from the nucleosome and can be covalently modified in many ways impacting chromatin structure (4). There are also variant forms of three of the four core histone proteins (H2A, H2B, and H3) that play different roles in cells (4). Many different histone modifications are possible, and some of these include acetylation, methylation, and phosphorylation (37). Through their modifications, histone proteins are able to recruit proteins to the nucleosome (37).

In addition to histone modifications, two other recognized epigenetic mechanisms include DNA methylation and the incorporation of variant histones. The most common DNA modification is methylation of the CpG dinucleotide (1). Methylation of DNA is seen as a repressive mark that results in recruitment of proteins and protein complexes that

maintain the heterochromatic, condensed chromatin state (1). The process of DNA methylation is accomplished through DNMTs, which are a family of enzymes that add methyl groups to DNA (1). Variant histones are incorporated into the nucleosome through the activity of histone chaperones or chromatin remodeling complexes (1). These variant forms provide a specific recognition surface for further recruitment of chromatin-associated complexes (1). These two mechanisms allow for recognition by proteins and protein complexes that can affect chromatin structure.

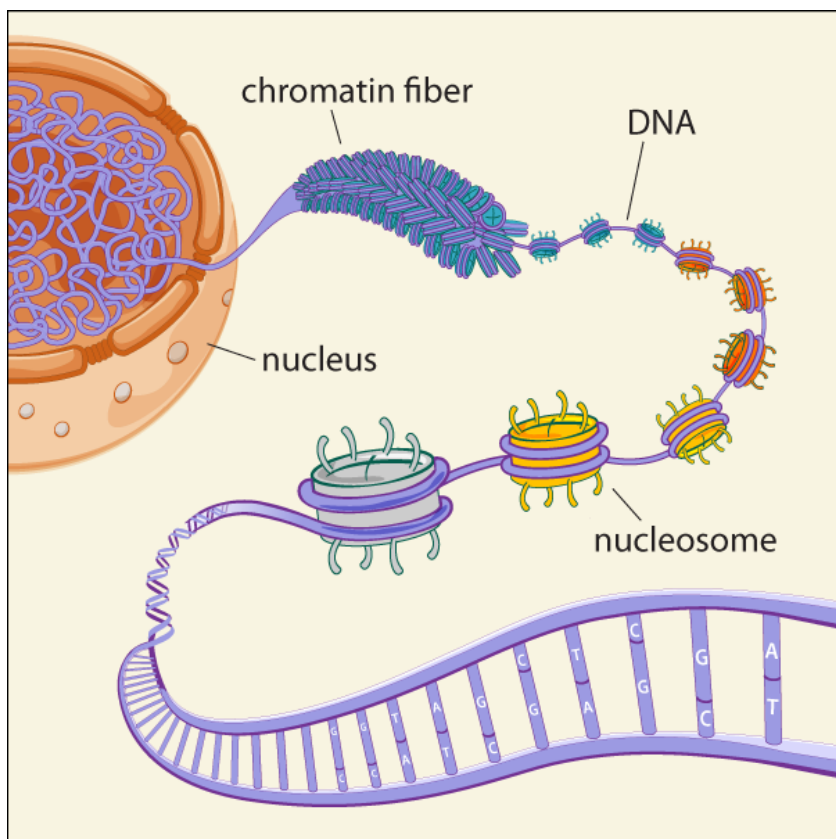


Figure 1. Cartoon Representation of Chromatin Compaction In the Nucleus of a Cell (13). The DNA (purple ladder) in a cell is wrapped around the histone octamer and generates the nucleosome (as indicated). This nucleoprotein complex is further condensed into the chromatin fiber and packaged into the nucleus of a cell.

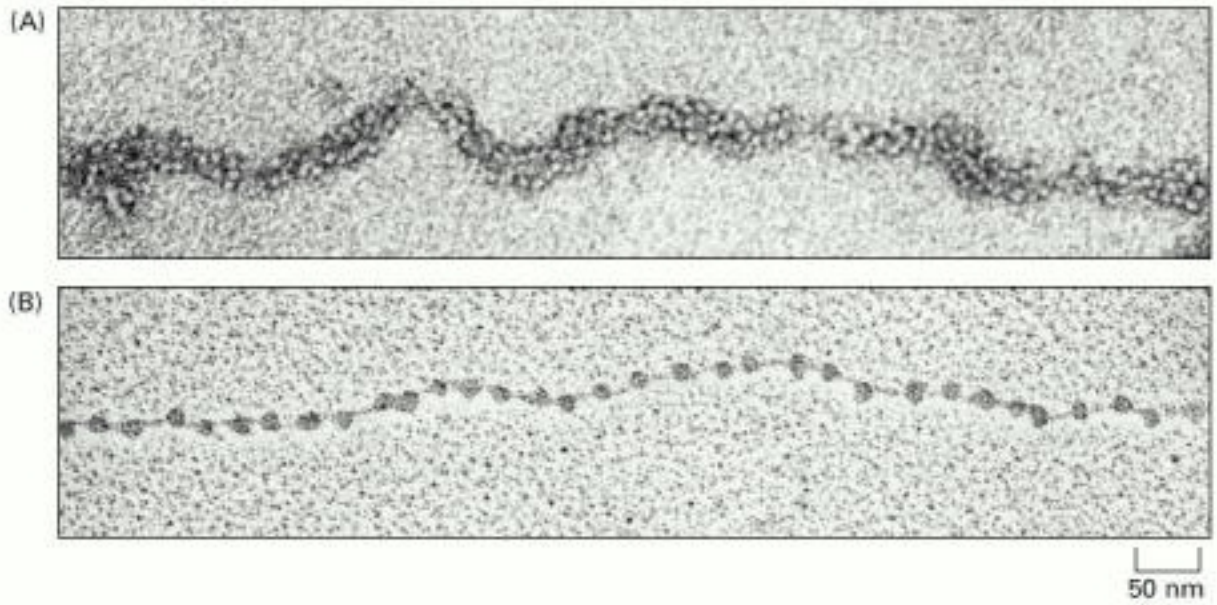


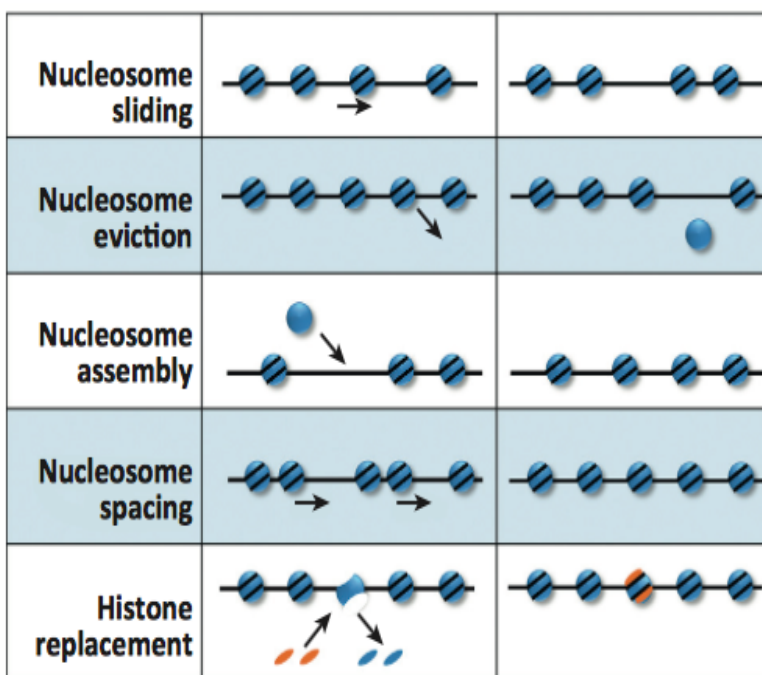
Figure 2. Electron Microscopy of Chromatin Fiber. A. 30nm chromatin fiber. B. Beads on a string model of chromatin (5). Figure taken from *Molecular Biology of the Cell* 4th edition (5).

Chromatin Remodeling

Chromatin remodeling complexes function to regulate chromatin structure through their abilities to alter the state of nucleosomes. These multi-protein complexes can replace histones with variant histones as well as slide, evict, or assemble nucleosomes (Figure 3) (32). Chromatin remodelers ultimately play a role in important cellular events such as gene regulation, DNA repair and replication, and recombination (32). These complexes use the energy derived from the hydrolysis of ATP to change the structure of nucleosomes (32).

There are four families of ATP dependent chromatin remodelers: SWI/SNF (Switch/Sucrose Non-Fermentable), ISWI (Imitation Switch), CHD (Chromodomain Helicase Binding Protein), and INO80 (Inositol Requiring Mutant 80). These complexes are grouped based upon the sequence homology of the ATPase subunit (1). The SWI/SNF family can function in both nucleosome sliding and eviction (32). Members of this family are involved in large complexes and play a role in many cellular functions such as embryonic stem cell differentiation and metabolism of lipids and glucose (10). The ISWI family aids in nucleosome spacing and assembly, and also plays a role in the organization of higher order chromatin structure (32). ISWI remodelers have also been shown to be involved in the nucleosome activity surrounding the DNA replication fork both before and after replication (18). It has also been shown that these complexes function in DNA repair through both interactions with nucleosomes as well as the recruitment of DNA repair proteins (18). CHD remodeling complexes have a variety of functions and are capable of nucleosome assembly and disassembly, as well as the sliding and spacing of nucleosomes (32). Members of the CHD family have been shown

to possess histone deacetylase activity in addition to chromatin remodeling (28). This family also plays a role in transcriptional elongation as well as development and differentiation (28). Finally, the INO80 family is involved in the replacement of canonical histone proteins in the nucleosome with their variant forms (32). Through the recognition of γ -H2AX, INO80 is able to function in DNA replication checkpoints as well as DNA damage repair (28). Altogether, these complexes function to alter the structure of chromatin through their abilities to change both the nucleosome composition and its position.



TRENDS in Genetics

Figure 3. Cartoon Representation of the Function of Chromatin Remodeling Complexes. (32). Demonstration of each of the functions of chromatin remodelers, the arrows represent the direction of movement or action. The nucleosomes are blue circles, and variant histones are represented in red.

These complexes can be recruited to specific sites in the genome through histone modifications, DNA sequences, and/or transcription factors (1). The combination of these features allows for specificity of interaction and may help define the different actions of these complexes. The subunits of these complexes have domains that specifically recognize histone modifications such as Plant Homeodomain (PHD) fingers, which frequently recognize methylated histones (1). Other domains may include chromo and bromo domains that recognize methylated and acetylated histones, respectively (1). Therefore, the interaction with these complexes and chromatin is likely complex involving many subunits and protein domains.

Nucleosome Remodeling Factor (NURF)

The nucleosome remodeling factor (NURF) is an ISWI family chromatin remodeler. NURF functions to slide nucleosomes in *cis* in an ATP-dependent manner (33). This action opens up chromatin structure and allows for gene expression to take place. The complex contains three subunits: BPTF, the largest and essential subunit, SNF2L (Sucrose Non-fermentable Protein 2L), the ATPase subunit, and pRBAP48, a WD repeat protein (Figure 4). The NURF complex is recruited to sites in the genome through interaction with both histone modifications and transcription factors (6). NURF has been previously shown to interact with Progesterone Receptor (PR), Smad, AP-1, SRF (Serum Response Factor), and CTCF (CCCTC Binding Factor) transcription factors (33). These interactions provide evidence that NURF may play a role in both cell-type specific and ubiquitous regulation of gene expression. It has been suggested through *in vitro* experiments that NURF is able to interact with transcription factors to move nucleosomes enabling transcriptional activation, and this suggests that NURF may be required for the regulation of transcription *in vivo* (7).

The NURF complex has been shown to play a role in many cellular processes including development and cell type specific gene activation through its ability to remodel nucleosomes and activate gene expression. In *Drosophila*, NURF plays a role in the maintenance of stem cells in the testis (15). It is proposed that this maintenance is due to the activation of the JAK/STAT pathway (15). In addition to this, NURF localizes with Ctf in three different mouse cell types (33). Ctf is an architectural protein that has a multitude of functions in gene regulation and can be found at enhancers and promoters as well as function as an insulator (33). The interaction of

NURF with Ctcf, suggests that NURF may play a role in regulating the cell-type specific binding of transcription factors to DNA near Ctcf sites (33).

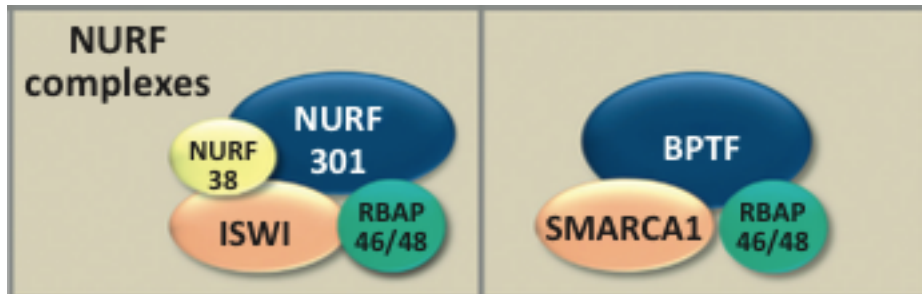


Figure 4. Cartoon representation of the *Drosophila* (left), and mammalian (right) NURF complexes (32). The subunits of the NURF complex are represented in each panel, and homologous domains are in the same colors.

BPTF

The Bromodomain PHD finger Transcription Factor, BPTF, is the largest subunit of the NURF complex, and is involved in the regulation of transcription (39). It is located on human chromosome 17q24.3 and the complete sequence encodes a protein with 2781 amino acids (39). This large protein contains many domains that allow it to function as a transcriptional regulator and play a role in chromatin remodeling. These domains can be seen in Figure 5, and consist of two PHD finger domains, Bromodomain, DDT (DNA binding homeobox and Different Transcription factors) domain, HMGA (High Mobility Group Protein A) domain, and a Q-rich domain (39). The C-terminal PHD finger recognizes H3K4me3, and the bromodomain recognizes H4 K12, K16, and K20 acetylation (39, 35). A previous study has suggested that the affinity of the bromodomain for H4K16Ac is enhanced when paired with the recognition of H3K4me3 by the C-PHD domain, and this multivalent interaction could enhance the recruitment of BPTF to chromatin (35). The HMGA domain interacts with nucleosomes through its acidic patch and AT-hook sequences (40). AT hook sequences are short sequences that bind to AT-rich DNA sequences in the minor groove of DNA (40). The DDT domain has been shown in our lab previously to play a role in holding the NURF complex together through interactions with the SNF2L subunit. The N-terminal PHD finger domain has not been previously characterized outside of its classification as a PHD finger domain. Together, these domains facilitate the interactions of BPTF with chromatin and transcription factors.

Through its role as a transcription factor, Bptf plays a role in many cellular processes and improper function has been indicated in both disease and malignancy. In

the mouse, Bptf expression is important for the development of the VE and DVE in embryos (23). This study also found that in Bptf knockout ES cells genes related to cell division, proliferation, and development are misregulated (23). A loss of Bptf results in embryonic lethality, indicating that Bptf plays an important role in the control of gene expression during early embryonic development (23). In addition to its role in development, Bptf has been suggested to play a role in a variety of cancer types. A common translocation seen in a variety of cancer types such as lung, liver, prostate, and breast cancers involves the partial gain of chromosome 17q (14). A non-reciprocal translocation found in human lung embryonal-derived cells was mapped to 17q24.3, which is the location of the BPTF gene (14). Cells with this translocation were characterized as having elevated proliferation *in vitro* and elevated levels of endogenous BPTF suggesting the translocation impacts the expression of endogenous BPTF (14). In the same study, it was shown that there is an abnormal copy number of the BPTF locus in lung cancer and neuroblastomas (14). It was also recently shown in melanoma, that knockdown of BPTF results in decreased cell proliferation and an upregulation of BPTF results in increased tumor growth (16). The downregulation of BPTF results in reduced ERK levels, suggesting that BPTF plays a role in MAPK signaling (16). In melanoma tissue samples, it was shown that there is an increased copy number of BPTF compared to normal tissue (16). This suggests that BPTF could be used as a predictor of metastasis in melanoma (16). Finally, BPTF may play a role in EMT in colorectal carcinoma (41). In this study, advanced tumor progression was correlated with elevated levels of BPTF, and that BPTF could be a regulator of EMT through the regulation of the expression of vimentin and E-cadherin (41). Altogether,

these findings indicate that BPTF plays a role in many important cellular processes as a transcriptional regulator and that aberrant expression and/or activity may result in disease or malignancy.

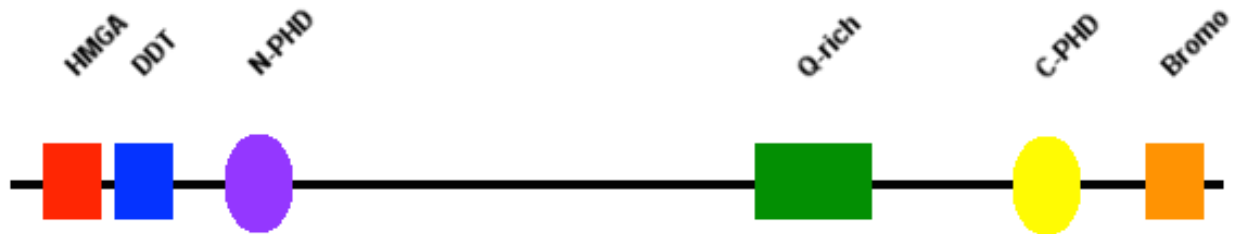


Figure 5. Functional domains of BPTF. The colored shapes represent the respective domains. All of these domains have been classified, with the exception of the N-PHD (N-terminal PHD finger) domain, which will be one of the focuses of this study.

PHD Finger Domains

The plant homeodomain (PHD) finger is a functional domain found in numerous nuclear proteins that plays a role in the regulation of chromatin structure (30). These domains commonly recognize histones and their modifications, while others are able to bind directly to DNA or RNA (30, 24). These domains have a characteristic Cys₄-His-Cys₃ motif that allows the coordination of two Zn²⁺ ions and they are around 60 amino acids in length (8). The interaction surface has been well characterized for the C-terminal PHD finger domain of Bptf and the PHD finger of ING2, both of which recognize H3K4me3 (8). In these proteins, a “cage” is formed around the histone modification, and it is hydrophobic or aromatic residues that form this structure and stabilize the interaction (8). This structure seems to be conserved across PHD finger domains (8).

Though PHD finger domains are classified based on their conserved motif, there does not seem to be one single shared ligand among them. Evidence has been shown that PHD finger domains specifically recognize; H3K4me3, non-modified histone tails, H3K9me3, acetylated H3 or H4, DNA, or non-histone proteins (25). As mentioned above, The C-terminal PHD finger and bromodomains of Bptf participate in chromatin binding through their abilities to recognize H3K4me3 and H4K16Ac, respectively. This pairing of a PHD finger domain with a nearby bromodomain is also seen in ACF1 and p300, proteins that have been shown to robustly interact with nucleosomes through histone modifications (11). In addition to histones, there is evidence that a subset of proteins with PHD finger domains may interact with specific protein ligands (11). One example of this is the protein Pygopus, found in *Drosophila*. This protein interacts with

Legless/BCL9 through its PHD finger domain (11). This interaction has been shown to occur specifically with the second cysteine residue of the conserved PHD finger motif and four residues in the second loop (11). There were three other amino acid residues important for this interaction that were not conserved among other PHD fingers, but they are conserved among all of the known Pygopus proteins (11). The binding of this protein to BCL9 results in transcriptional activation through interaction with β -catenin (11). In addition to these interactions, it was shown for the first time in 2012, that a PHD finger domain binds to DNA (25). It was determined that one of the PHD finger domains of BRPF2 binds DNA in a non-sequence specific manner (25). This PHD finger domain possessed a patch of basic residues not found in other typical PHD finger domains, while it still maintained the signature Cys₄-His-Cys₃ motif (25). PHD finger domains are still under study, and while they share the characteristic motif, they may not be that similar in their interaction partner despite having similar functions in cellular metabolism.

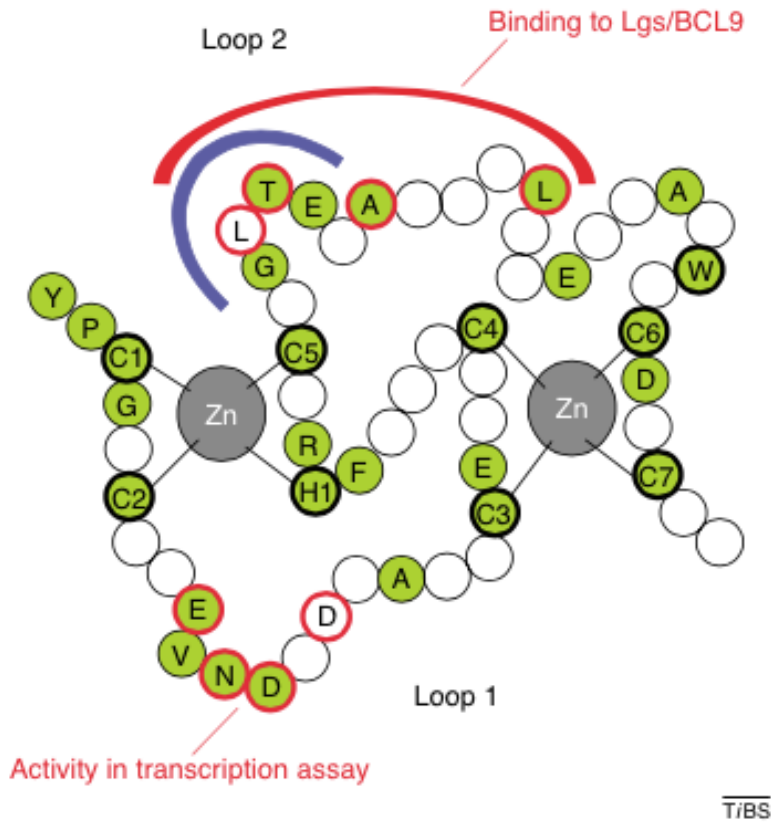


Figure 6. Spatial representation of the PHD finger domain in Pygopus (11). The characteristic Cys₄-His-Cys₃ residues are outlined in black, and it can be seen how this peptide folds to hold two Zn²⁺ ions. Other important residues for this particular protein are highlighted in red, and the residues important in binding the protein partner, BCL9 are shown with the red arch.

Rationale for Study:

Our lab currently has genome wide data on the localization of NURF to the genome. Through this data, it is suggested that NURF is found at DNaseI hypersensitivity sites, Ctf binding sites, and at both the middle and ends of genes. In an effort to piece this information together, my study will provide further insight into the molecular mechanisms through which NURF is specifically recruited to sites in the genome through the investigation of the N-terminal PHD finger domain (N-PHD). It was of interest to determine the binding partner for this domain in an effort to further understand the function of Bptf, and the NURF complex as a whole. In order to do this, pull down experiments were performed to test the interaction of this domain with histones, specific histone modifications taken from the genome wide data, and nucleic acids. Protein partners were identified through the generation of a column with N-PHD covalently linked to resin, and nuclear extract poured over the column to pull down potential interaction partners. Protein bands identified from this experiment were then identified through mass spectrometry. I also worked on expressing a cDNA to Bptf in order to test NURF recruitment to specific sites in the genome through a series of mutations in its known domains. A cDNA for Bptf was previously generated in the lab and found to not express in mouse ES cells. A sequence representing two exons was cloned into the Bptf cDNA. The addition of this sequence was then tested in mouse ES cells using nucleofection, protein extraction, and western blotting for the streptavidin tag on the cDNA. The cDNA is important to dissect the interaction of Bptf, and the NURF complex as a whole with specific sites in the genome to further understand its recruitment.

Specific Aims:**Aim 1:**

Clone two exons into a Bptf cDNA to test the expression of exogenous Bptf in mouse ES cells.

Hypothesis: Insertion of the two exons missing from the Bptf cDNA will allow for the exogenous expression of Bptf in mouse ES cells.

Aim 2:

Identify the binding partner for the N-terminal PHD finger domain (N-PHD) of Bptf.

Hypothesis: N-PHD of Bptf has a binding partner, and this binding partner is likely a chromatin-associated protein.

Chapter 2: Materials and Methods

Cell Culture

P19, mouse embryonic carcinoma cells, were grown in Dulbecco's modified Eagle medium (DMEM) with 10% FBS, 1X penicillin and streptomycin, 2 mM glutamine, and 1X non essential amino acids. The cells were then cultured until they reached >85% confluence. At this stage, the cells were harvested from the plates. The media was removed, and the cells scraped using a cell scraper. The cells from ten plates (~700 million cells) were collected in one 15 mL conical tube, and spun at 2000 rpm for 5 minutes. Fresh culture medium was added back to the plates to maintain a continuous culture of cells. The resulting pellet was washed with 1X PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄ pH 7.4) to remove any residual culture medium. The washed pellet was then flash frozen in liquid nitrogen and stored at -80°C until use.

CJ7, mouse embryonic stem cells, were grown on gelatinized plates. They were grown in DMEM with 15% ESC grade FBS, 1X penicillin and streptomycin, 2 mM glutamine, 1X nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 1000 units/mL leukemia inhibitory factor (Millipore). The cells were plated on one 10 cm² dish, and maintained in culture until ready for nucleofection.

Cloning

To clone the N-terminal and C-terminal PHD finger domains of Bptf, primers were designed flanking each domain as shown in Table 1. The full-length Bptf was used as the template at 50ng per reaction using Phusion Polymerase (NEB). The PCR protocol used was as follows: 98°C 30 seconds, 40 cycles of 98°C 10 seconds, 68°C 30 seconds, 72°C 30 seconds, and a final extension at 72°C for 5 minutes. Following amplification, the PCR product was purified using a PCR Cleanup kit (Qiagen), and digested using BamHI (NEB) and XhoI (NEB) in a 37°C water bath overnight. The insert was then ligated into the pGEX4T-1 GST fusion vector, and transformed into GC10 *E. coli*. Clones were confirmed through restriction digest with BamHI and XhoI and transformed into BL21-CodonPlus(DE3)-RIPL strain *E. coli* (Agilent Technologies) using ampicillin and chloramphenicol for selection.

To clone the additional exons for insertion into the Bptf cDNA, primers were chosen that flanked the region as shown in Table 1. The general scheme for the cloning of the exons is shown in Figure 7. cDNA from mouse embryonic fibroblasts (MEF) was used as the template at 50 ng per reaction using Phusion Polymerase. The PCR protocol was done as follows: 98°C 30 seconds, 30 cycles of 98°C 10 seconds, 55°C 30 seconds, 72°C 60 seconds, and a final extension at 72°C for 5 minutes. The resulting 700 bp band was cloned into the TOPO PCR Blunt vector. The Bptf cDNA was digested with BspEI and partially digested with NheI at 37°C overnight and the resulting 3 kb fragment cloned into the pACYC184 vector. The 700 bp band was extracted from the TOPO vector using HpaI and NdeI and cloned into the pACYC184 vector containing the Bptf fragment. The resulting clones were screened and confirmed via Sanger

Sequencing done at the VCU Sequencing Core Facility. The expanded insert was then put back into the original Bptf cDNA using BspEI and NheI. Clones were confirmed via Sanger Sequencing at the VCU Sequencing Core Facility.

To generate MBP fusion proteins the primers used are as outlined in Table 1 for SRSF9, Hmgb1, and Thoc4. Mouse cDNA from CJ7 cells was used as the template at 50 ng per reaction and the PCR protocol was as follows: 98°C 30 seconds, 30 cycles of 98°C 10 seconds, 45-65°C gradient 30 seconds, 72°C 30 seconds, and a final extension at 72°C for 5 minutes. Amplification was performed using Phusion Polymerase. The resulting PCR product was run on a 0.8% agarose gel, and lanes containing product were extracted using Gel Purification Kit (Qiagen). The purified insert was then cloned into the pMAL c2X MBP fusion protein vector and transformed into GC10 cells. Clones were confirmed and transformed into BL21-CodonPlus(DE3)-RIPL *E. coli* (Agilent Technologies) using ampicillin and chloramphenicol for selection.

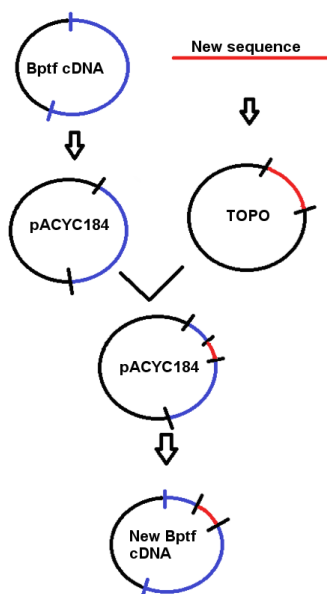


Figure 7. Schematic of Cloning Strategy for Bptf cDNA. Blue represents Bptf sequence, black represents plasmid backbone, and red represents additional sequence cloned.

Recombinant Protein Expression and Purification

To express recombinant proteins (GST and MBP fusion proteins) transformed BL21 for each construct were used to inoculate 5 mL of LB with ampicillin (0.1 mg/mL) and chloramphenicol (0.034 mg/mL) and grown overnight at 37°C with 250 rpm shaking. The next day, 1 mL of the 5 mL starter culture was used to inoculate 100 mL LB with ampicillin (0.1 mg/mL) and chloramphenicol (0.034 mg/mL). The culture was grown at 37°C for around 2 hours, or until the OD₆₀₀ was around 1.0. The culture was induced with 0.5 mM IPTG and 2% ethanol, and grown at 18°C overnight with 250 rpm shaking. The following day the culture was pelleted at 5000 rpm for 5 minutes. The pellet was stored at -20°C.

For the GST fusion proteins, the pelleted *E. coli* were resuspended in 1/10 culture volume PBS, 1% Triton X-100 with 1X protease inhibitor cocktail without EDTA (Roche). The resulting suspension was sonicated for 3 minutes total, 30 seconds on, 2 minutes off, at 4.0 output. The lysate was then spun at 10,000 Xg for 10 minutes to remove the insoluble fraction. The resulting soluble fraction was transferred to a new tube. Glutathione resin (glutathione reduced, immobilized on Agarose CL-4B, GE Life Sciences) was washed with PBS, 1% Triton X-100 to remove storage buffer. The resin was then added to the soluble fraction and mixed continuously at 4°C for two hours. Unbound protein was washed away with ten resin volumes of PBS. The resin was then doubled in volume with glycerol to make a 50% glycerol stock, and stored at -20°C.

The MBP fusion proteins were generated in a similar manner. The pelleted *E. coli* were resuspended in 1/10 volume PBS, 1% Triton X-100 and 1X protease inhibitor cocktail without EDTA (Roche). The soluble fraction was generated in the same way as

described above. The MBP fusion proteins were pulled out of solution using Amylose resin (New England Bio Labs) that was pre-washed with 1X PBS 1% Triton X-100. The proteins were pulled down with continuous mixing at 4°C for two hours in a 10 mL disposable column. The unbound protein was washed with ten resin volumes of PBS. For the immobilized proteins, glycerol was added to 50% and stored at -20°C. Bound protein was eluted from the column using 10 mM maltose to generate purified protein to be used in *in vitro* pull downs. The resulting protein was then dialyzed using 3.5 kDa cutoff dialysis tubing against PBS to remove the maltose, and then quantified using the BioRad DC Protein assay (BioRad). The protein was then diluted to 100 ng/μL, aliquoted, and stored at -80°C until use.

Pull Downs

1. Nucleic Acid Pull Downs

Equal loading of immobilized GST fusion N-PHD and C-PHD proteins as well as immobilized GST (MBP fusion proteins included for RNA Pull Down) were washed with Nucleic Acid Pull Down Buffer (NAPDB) (25 mM Hepes pH 7.6, 100 mM NaCl, 5 mM MgCl₂, 3 mM CaCl₂, 10% glycerol, 0.1% NP40). To the washed resin, 500 μL of NAPDB was added followed by 20 μg of sonicated salmon sperm DNA (10 mg/ml) or 800 ng mRNA. The resulting mixture was incubated on an end-over-end rotator at 4°C for three hours. Following the incubation, unbound DNA or RNA was washed from the resin using NAPDB, and eluted in 20 μL 0.1% SDS in TE buffer and heated at 65°C for 30 minutes. The resulting sample was run on a 1% agarose gel with 5 μL 6X Orange G DNA loading dye. Input was prepared as follows; 90 μL TE, 10 μL 6X Orange G DNA loading dye, and 10 μg sonicated salmon sperm DNA or 400 ng RNA.

2. Histone Pull Down

Equal loading of immobilized GST fusion proteins (C-PHD, N-PHD, GST) were washed with Pull Down Buffer (PDB) (25 mM Hepes pH 7.6, 100 mM NaCl, 5 mM MgCl₂, 3 mM CaCl₂, 10% glycerol, 0.1% NP40). One CJ7 cell pellet (~70 million cells) was resuspended in MNase digestion buffer (25 mM Hepes pH 7.6, 100 mM NaCl, 5 mM MgCl₂, 3 mM CaCl₂, 10% glycerol, 0.1% NP40, 2 uM pepstatin, 100 uM Leupeptin, 5 uM phosphoramidon, 10 uM MG132, 1X phosphatase inhibitor, 1000 units MNase, 1X protease inhibitor cocktail without EDTA (Roche)). The cells were lysed via Dounce homogenization, and kept on ice for two hours to allow complete digestion of chromatin. The lysate was spun down for 15 minutes at 10,000 x g, and the resulting supernatant was saved and adjusted to 150 mM NaCl using 5 M NaCl. The resulting solution was split evenly among the resin, saving equal volume for input, and incubated at 4°C on an end-over-end rotator for three hours. Following incubation, the resin was washed with PDB, and the protein eluted with 40 µL 2X SDS loading buffer at 65°C for 30 minutes. The proteins were run on a 15% polyacrylamide gel 1 hour 200V, 200 mA in 1X SDS buffer (25 mM Tris, 192 mM glycine, 1% SDS) and transferred to PVDF membrane (BioRad) 17 hours 20 V, 20 mA in 12.5 mM Tris, 95 mM glycine, 0.1% SDS, 20% methanol. The membrane was blocked with 5% NFDM in PBST (PBS, 0.1% Tween-20) for 2 hours and probed with H3 primary antibody (Abcam, 1:1000 dilution, ab1791) overnight at 4°C. The primary antibody was washed off with PBST, and the membrane was probed with the secondary antibody α-rabbit HRP (Cell Signaling cat#70745, 1:10000 dilution) for 1 hour at room temperature. The unbound antibody was washed with PBST, and the membrane treated with Femto ECL reagent (ThermoScientific

SuperSignal West Femto Luminol/Enhancer Solution, cat#1859022 and Stable Peroxide Buffer, cat#1859023) and exposed on film to identify proteins.

3. Peptide Pull Down

Biotinylated peptide sequences were bound to streptavidin beads at 200 pmole final concentration. Briefly, 10 μ L streptavidin bead slurry (Invitrogen – Dynabeads M-270 Streptavidin cat#65306) was used for each pull down. The beads were washed with 1X PBS, and binding was done in PBS overnight at 4°C on an end-over-end rotator. The next day, the beads were washed with Peptide Pull Down Buffer (PPDB) (20 mM Hepes pH 7.6, 150 mM KCl, 0.2% Triton X-100) and resuspended in an appropriate volume to use 10 μ L 50% slurry per pull down. To each pull down, 1 μ g purified GST or GST fusion protein was added with PPDB and binding was done on ice for 1 hour with frequent mixing. Unbound protein was washed with PPDB followed by one wash with 1X PBS to remove the KCl. Bound proteins were eluted with 1% SDS at 65°C for 30 minutes. The supernatant was removed and 20 μ L 5X SDS protein loading buffer added and heated an additional 15 minutes at 65°C. The pull down samples were run on a 12% Tris-Glycine acrylamide gel for 1 hour 200 V, 200 mA, and transferred to PVDF membrane for 1 hour at 100 V, 350 mA in 1X Tris-Glycine 0.1% SDS transfer buffer. The membrane was blocked with 5% NFDM in PBST, and probed with primary antibody against GST (Cell Signaling, 1:1000 dilution, cat#2625) overnight at 4°C with gentle agitation. The membrane was washed with PBST and probed with secondary antibody to α -rabbit HRP (Cell Signaling, 1:10000 dilution, cat#70745) for one hour at room temperature with gentle mixing. Finally, the membrane was washed with PBST and exposed using Femto ECL reagent and exposed to film.

Affinity Column

1. Protein Expression

GST fused N-PHD and GST expressed in BL21 *E. coli* were grown in a starter culture of 100 ml LB with chloramphenicol (0.034 mg/mL) and ampicillin (0.1 mg/mL) overnight at 37°C with 250rpm shaking. The following day, the 100 ml culture was added to 3 L LB with chloramphenicol (0.034 mg/mL) and ampicillin (0.1 mg/mL) and grown at 37°C with 100 rpm until OD₆₀₀ was 1.0. Protein expression was induced with 0.5 mM IPTG and 2% ethanol, incubated at 18°C overnight with 100 rpm shaking, and pelleted via centrifugation at 5000 rpm for 5 minutes. The pellet was stored at -80°C until needed.

2. Protein Purification

The 3 L pellet was resuspended in 30 mL PBS, 1% Triton X-100 plus 1X protease inhibitor cocktail without EDTA (Roche). The resulting suspension was sonicated 5 minutes total, 30 seconds on, 2 minutes off, 4.5 output and spun down at 10000 xg for 15 minutes. The soluble fraction was saved and poured over a 2 mL glutathione-agarose resin column twice. The column was washed with 10mL PBS, 1% Triton X-100, and the bound protein eluted with 10mM glutathione reduced in PBS, 1% Triton X-100. The eluted protein was dialyzed against Coupling Buffer (0.1 M sodium citrate, 0.05 M sodium bicarbonate pH 10.0), using 3.5 kDa cutoff dialysis tubing, for two days. The resulting protein was concentrated to ~1mL using Amicon Ultra centrifugal filters at 10 kDa membrane cutoff. The final protein concentration was calculated using the BioRad DC Protein Assay (described below).

3. Protein Coupling

To generate the affinity column, the AminoLink resin (Thermo Scientific) was washed with Coupling Buffer to remove residual storage buffer. 10 mg of both GST and GST N-PHD were added to 1mL AminoLink resin and allowed to bind overnight with end over end mixing at room temperature. The next day, the unbound protein was washed away with PBS. The bound protein was coupled with 50 mM NaCNBH₃ in PBS for four hours at room temperature with end over end mixing. The solution was removed and the column washed with Quenching Buffer (1 M Tris HCl pH 7.4, 0.05% NaN₃) and the remaining active sites were blocked with 50 mM NaCNBH₃ in Quenching Buffer. Finally, the column was washed with Wash Solution (1 M NaCl, 0.05% NaN₃). The column was stored at 4°C in Preservation Buffer (25 mM Hepes pH 7.6, 150 mM NaCl, 0.1% NP40, 2 mM MgCl₂, 10% glycerol, 0.05% sodium azide) until use. Schematic of the columns generated is shown in Figure 8.

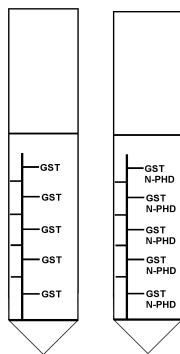


Figure 8. Schematic of the two Affinity Columns Generated. Each column has GST or GST N-PHD covalently linked to the resin. A total of 1 mL of resin makes each column. These columns were used to analyze the binding partner of N-PHD through the use of P19 nuclear extract poured over the column and the resulting protein fractions were separated by SDS-PAGE and unique protein bands analyzed by mass spectrometry.

4. Nuclear Extract

A nuclear extract was prepared using a high salt extraction protocol (17). P19 mouse embryonic carcinoma cells were used to make the extract. Fifty 10 cm² dishes (~3.5 billion cells) were used to generate the cell pellet (harvested as described above). The pellet was thawed and rinsed in 10 mL Buffer A (10 mM Hepes pH 7.6, 1.5 mM MgCl₂, 10 mM NaCl). The pellet was then resuspended in 10 mL Buffer A plus 1X protease inhibitor cocktail without EDTA (Roche). The cells were lysed with 15 strokes using a 5 mL dounce homogenizer, and kept on ice for 10 minutes. This process was repeated for a total of two times. The suspension was centrifuged for 10 minutes at 2000 rpm to pellet the nuclei. The supernatant was discarded and the nuclei pellet was resuspended in 5 mL Buffer C (20 mM Hepes pH 7.6, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl₂) plus 1X protease inhibitor cocktail without EDTA (Roche). The nuclei were lysed with 15 strokes using a 5 mL dounce homogenizer, kept on ice for 10 minutes, and this step was repeated once. The solution was spun for 10 minutes, max speed, 4°C in a standard tabletop centrifuge. The soluble fraction was saved and the NaCl concentration adjusted to 300 mM using Buffer D (20 mM Hepes pH 7.6, 20% glycerol, 0.2% Triton X-100) and 1X protease inhibitor cocktail without EDTA (Roche). The extract was kept on ice for 30 minutes, mixing frequently, and spun once more for 10 minutes at 10000g, at 4°C, to remove any insoluble protein. The extract was then split in two equal volumes.

5. Pull Down

The affinity column with both GST and GST N-PHD was washed with Column Buffer (25 mM Hepes pH 7.6, 300 mM NaCl, 0.1% NP40, 2 mM MgCl₂, 10% glycerol)

followed by a wash with 1 mM ZnCl_2 in Column Buffer, and finally another wash with Column Buffer to remove residual ZnCl_2 . The zinc wash ensures the PHD finger is properly folded. The nuclear extract from above was poured over each column twice, and then the column was washed with 10 column volumes of Column Buffer. The bound proteins were eluted in three fractions (3 column volumes each):

- a) 1 mM ATP, 25 mM Hepes pH 7.6, 300 mM NaCl, 0.1% NP40, 2 mM MgCl_2 , 10% glycerol
- b) 500 mM NaCl, 25 mM Hepes pH 7.6, 0.1% NP40, 2 mM MgCl_2 , 10% glycerol
- c) 1M NaCl, 25 mM Hepes pH 7.6, 0.1% NP40, 2 mM MgCl_2 , 10% glycerol

The column was then washed with five column volumes, 25 mM Hepes pH 7.6, 1 M NaCl, 0.1% NP40, 2 mM MgCl_2 , 10% glycerol, five column volumes 25 mM Hepes pH 7.6, 300 mM NaCl, 0.1% NP40, 2 mM MgCl_2 , 10% Glycerol, and finally three column volumes 25 mM Hepes pH 7.6, 150 mM NaCl, 0.1% NP40, 2 mM MgCl_2 , 10% glycerol, 0.05% sodium azide. The column is stored in the preservation buffer wash solution at 4°C. The protein fractions were precipitated with 20% trichloroacetic acid, and stored on ice overnight. The samples were spun 10 minutes at 21,000 xg to pellet the protein. The resulting pellets were washed with acetone four times with shaking at each wash. The pellet was dissolved in 45 μL 1X SDS protein loading buffer in 8M urea, 1% SDS and heated at 65°C for 30 minutes or until pellet dissolved completely. The samples were then run on either a 4-20% Tris-glycine acrylamide gradient gel or a 15% Tris-glycine acrylamide gel and silver stained using BioRad Silver Stain Plus kit. The gel was analyzed for differences between the experiment (GST N-PHD) and the control (GST) lanes. Bands identified as different were analyzed via mass spectrometry.

6. Mass Spectrometry Analysis

To identify the proteins pulled down in the column experiment, the bands were excised from the gel and analyzed via mass spectrometry. Briefly, the gel band was excised from the experiment lane and a matching area was removed from the control. The gel piece was cut into ~1 mm cubes, dehydrated via SpeedVac centrifugation, and then rehydrated with 12.5 ng/μL trypsin (Promega) in 100 mM AB 1% ProteaseMAX surfactant (Promega) on ice and then warmed to room temperature at which point the trypsin digested the protein in the gel. Trypsin cleaves the carboxy end of lysine and arginine, and therefore will generate peptide sequences with predictable size and composition for each protein. To extract the digested peptides, the gel piece is repeatedly dehydrated/rehydrated and the supernatant collected and pooled at each step. The resulting peptide pool is then dried via SpeedVac centrifugation and resuspended in 100 mM AB. The sample was loaded onto a self-packed fused silica (Polymicro Technologies) trap column (360 μm outer diameter X 100 μm inner diameter) with a Kasil frit packed with 5-15 micron irregular phynyl C-18 YMC packing. The trap column was connected to an analytical column (360 micron X 50 micron) with a fritted tip at 5 micron or less (New Objective) packed with 5 μm phenyl C-18 YMC packing. The peptides were trapped and eluted into a Thermo Finnigan LCQ deca XP max mass spectrometer with an acetonitrile gradient from 0-80% over one hour with flow rate between 50-150 nL/minute. The mass spectrometer was operated in data dependent mode. First a MS scan from mass 300-1600 m/z was collected to determine the mass of peptides eluting at that time, the top five most abundant masses were fragmented into MS/MS scans and placed on an exclusion list. This was repeated

through the hour gradient to generate approximately 5000 MS² scans. The scans were then searched on Sequest using a mouse non-redundant database downloaded from NCBI. The following variable modifications were considered: oxidized M XCorr cut off of (1.25, 1.75, 2.25) for (+1, +2, +3) peptide charge states. MS² scans passing this cut off were manually verified, and proteins were accepted as present with two or more peptides accepted with high confidence.

Nucleofection

To test the expression of the Bptf cDNA in mouse ES cells, a nucleofection was performed using the Mouse ES Cell Nucleofector Kit from Amaxa/Lonza. CJ7 mouse ES cells were grown on gelatin-coated plates, and the media was changed 12 hours before nucleofection. On the day of nucleofection, the cells were trypsinized, counted, and spun down in a separate tube for each DNA sample. 2.5×10^6 cells were used in each sample. 3.5 µg of DNA and 2 µg of GFP vector were added to 10 µL of Mouse ES Cell Nucleofector Solution. The ES cells were resuspended in 90 µL of Mouse ES Cell Nucleofector Solution. The cells and DNA were mixed gently by pipetting up and down and then transferred to a cuvette. The nucleofection was performed in a Nucleofector® I machine, and 500 µL pre-warmed culture medium was added immediately after. The cells were finally transferred to gelatin-coated dishes and cultured for 72 hours. 24 hours after nucleofection, expression of GFP was assessed. After 72 hours, the cells were treated with Tri-Reagent (Sigma) and the protein was extracted as described below. Proteins were quantified using DC Protein Assay (BioRad) as described below. To determine exogenous Bptf expression levels, equal amounts of each protein sample were run on a 4% Tris-glycine acrylamide gel and transferred to PVDF membrane

(BioRad) 20 V, 20 mA for 17 hours. The membrane was blocked with 5% NFDM and probed with Streptavidin-HRP antibody (Invitrogen, 1:10000 dilution, cat#19534-050) at 4°C overnight. The membrane was then washed with PBST and exposed using Femto ECL reagent to film.

Protein Extraction

To extract the protein from the transfected CJ7 cells, the media was removed and the cells washed with 1X PBS to remove excess culture medium. 1 mL of TRI Reagent (Sigma) was added per 10 cm² culture dish. The lysate was homogenized by pipetting up and down many times, and the resulting lysate transferred to a 1.5 mL tube. 200 µL chloroform was added to each sample and then vortexed for 15 seconds. Samples were incubated for 15 minutes at room temperature followed by a 15 minute spin at 12,000 xg at 4°C. This step results in a separation into three phases, pink organic phase (protein) on bottom, middle interphase (DNA), and upper aqueous phase (RNA). The two upper layers were removed and the proteins were precipitated with 1mL isopropanol, samples vortexed, and incubated at room temperature for 10 minutes. The samples were then spun at 12,000 xg for 10 minutes. The resulting pellet was then washed overnight at 4°C with continuous mixing in 1.5 mL 0.3 M guanidine in 95% ethanol. The next day, the samples were spun 12,000 xg for 10 minutes and the pellet was washed in 100% ethanol for six hours at 4°C with continuous mixing. Finally, the samples were spun max speed for 10 minutes in a standard tabletop centrifuge at 4°C. The pellet was then dissolved in 100 µL 8M urea, 1% SDS at 65°C overnight.

Protein Quantification

For protein quantification, the BioRad DC protein assay was used with serial

dilutions of BSA in PBS. 1 μ L of each standard and each unknown was used (unknown was diluted 1:10 if above the standard curve) 100 μ L Reagent A was added to each tube and mixed well by pipetting up and down. Then 800 μ L of Reagent B was added to each sample and incubated for 20 minutes at room temperature. The absorbance of each sample was then measured at 750nm on a spectrophotometer and input into an excel spreadsheet to determine the slope of the standard curve. The slope was then used to calculate the protein concentration of each unknown. For protein samples derived from Tri-Reagent, the samples were diluted in 8M Urea, 1% SDS and 5X SDS Running Buffer to make a 2 mg/mL stock.

mRNA Purification

mRNA was purified from total RNA extracted from CJ7 cells using Dynabeads Oligo (dT)₂₅ resin (Life Technologies). A total of 1 mL 50% slurry was used to pull down the mRNA from total RNA. The resin was washed with Binding Buffer (20 mM Tris-HCl pH 7.5, 1 M LiCl, 2 mM EDTA) and resuspended in 500 μ L Binding Buffer. The volume of total RNA was brought up to 1 mL with H₂O and 1mL Binding Buffer was added to make 2 mL total volume. The sample was then heated at 65°C for two minutes to destroy secondary structures. It was then placed on ice and the 2 mL of total RNA was added to the 1 mL of resin, and the solution was mixed continuously for five minutes at room temperature. A magnet was used to pull the beads to the side of the tube, and the supernatant removed. The resin was then washed with Washing Buffer (10 mM Tris-HCl pH 7.5, 0.15 M LiCl, 1 mM EDTA). The bound mRNA was eluted using 100 μ L 10 mM Tris-HCl at 75-80°C for two minutes, and the beads were then quickly pulled to the side

with a magnet and the supernatant collected in a new tube. The mRNA was then quantified, and stored at -80°C until used.

Table 1. Primers used in generation of GST/MBP fusion proteins, as well as the amplification of new sequence in Bptf. Each primer contains a restriction enzyme site to allow for easy cloning into the vector of choice.

Protein	Forward	Reverse
Bptf exons	CTGGGAAGCAGAACTGTGCAGA GT	GGTGCTGGTGGTTAACTCCC
N-PHD	GGGGGATCCGAAGGGGTAATCC AGTAT	GGGCTCGAGTCACACACCAGGC ACCTTGTG
C-PHD	GGGGGATCCAAGGAGGCCAAGA AGGAC	GGGCTCGAGTCAGGCGTCCTCT GTCGACTG
Srsf9	GGGGATATCATGTCGTCGGGCT GG	GGGGGATCCTCATCAGTAGGGC CTGAAAGG

Chapter 3: Results

Addition of two Exons Does not Enhance the Exogenous Expression of Bptf cDNA in Mouse ES Cells.

In order to examine the role of Bptf in the recruitment of NURF to specific sites of the genome, a cDNA for Bptf was generated (by a previous student in the lab, data not shown). However the cDNA generated was unable to be expressed in mouse ES cells. It is important to have this cDNA expressed in ES cells as the lab has genome wide data for NURF that was generated in mouse ES cells, and the goal is to be able to test the effect of mutations in the domains of Bptf on the localization of NURF to specific sites in the genome identified from this data.

An alignment of the human BPTF sequence and our cDNA sequence indicated that there were two exons missing from the cDNA that are found in human BPTF. Also included in this analysis was FAC1, which was initially discovered in the plaques of Alzheimer's patients. It was later discovered that this protein represents a shorter version of BPTF; FAC1 codes for the first 801 N-terminal amino acids of BPTF (19). It is suggested that FAC1 acts as a transcription factor through DNA binding. From this alignment it was shown that our cDNA was lacking a region spanning around 200 bp that corresponds to two exons found in both the human BPTF and FAC1 proteins (these exons are present in the mouse Bptf genomic locus). Therefore, this region was cloned

into the cDNA sequence and Bptf expression was tested in mouse ES cells. The new cDNA with the additional sequence was transfected into CJ7 cells (mouse ES cells) via electroporation using GFP as a positive control, and as seen in Figure 9 the GFP expression indicates that the DNA was successfully inserted into the cells. It was found that the addition of sequence did not increase expression of Bptf in ES cells (Figure 10). As a positive control, 293T cells were transfected with the cDNA and run alongside the protein extracts from ES cells (Figure 10). This suggests that there is some mechanism in the ES cells that is preventing the exogenous expression of Bptf from the cDNA. There is endogenous expression of Bptf in these cells, however in order to test mutations in the different domains of Bptf, the cDNA will provide an important tool in further exploring the recruitment of the NURF complex to the genome.

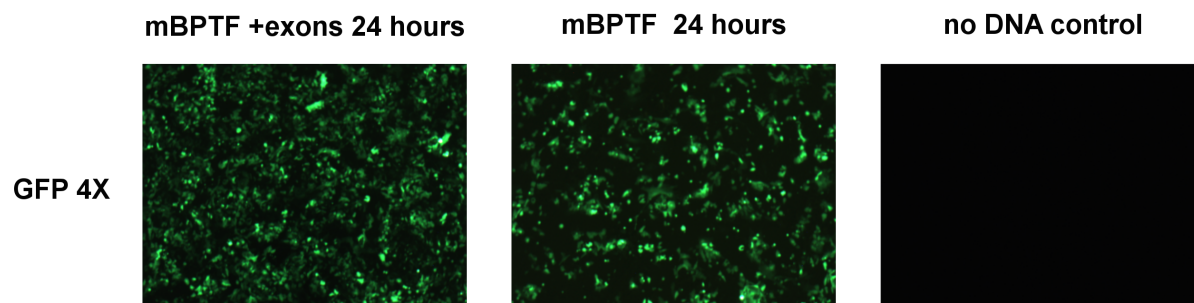


Figure 9. Visual of GFP expression in CJ7 cells. Photos were taken 24 hours after nucleofection. The negative control, no DNA, is completely black while the experimental samples show GFP expression for both cDNA constructs. Both the cDNA and GFP are under the expression of a CMV promoter.

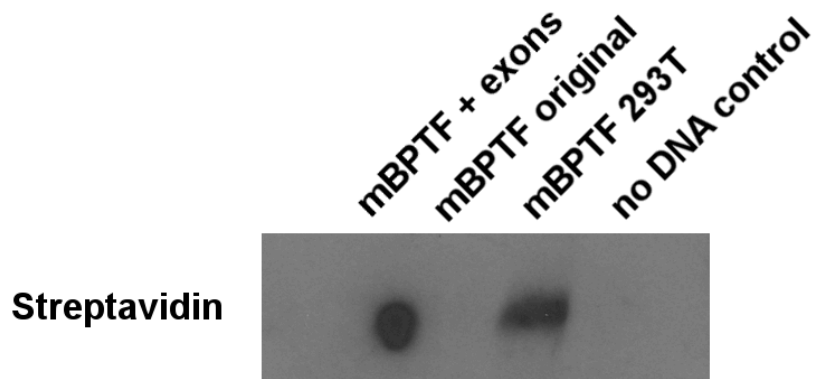


Figure 10. The Bptf cDNA is not Expressed in ES Cells. Protein levels as assessed by Streptavidin-HRP Western blot from protein extracts of CJ7 cells (mBPTF+exons, mBPTF original, and no DNA control) and 293T cells (mBPTF 293T). The addition of the two exons to the Bptf cDNA does not allow for the exogenous expression of the Bptf cDNA in mouse ES cells. The dot is an artifact from the piece of film.

The N-terminal PHD finger Domain of Bptf Does not Interact with Histones.

The first step in investigating the function of the N-PHD of Bptf was to investigate its interaction with histones. PHD finger domains commonly recognize histones in either their unmodified or modified states. In order to test the ability of N-PHD to bind histones, GST fusion proteins for both of Bptf's PHD finger domains were created.

To test whether this domain binds to known histone modifications, a peptide pull down experiment was performed with immobilized modified histone peptides and purified recombinant GST fusion proteins. These modifications were selected based upon genome wide data in the lab that shows overlap of NURF localization with H3K4me1, H3K4me3, and K3K36me3. Figure 11 shows that C-PHD interacts with H3K4me3, as has been previously reported (39). In this experiment it was also seen that N-PHD does not interact with any of the chosen histone modifications (Figure 11).

Based on these results, it was of interest to test the interaction with all histone proteins and their variants, a pull down was done with the GST fusion proteins and mononucleosomes. Mononucleosomes were generated from CJ7 cells via MNase digestion. The resulting mononucleosomes were then incubated with the GST fusion proteins, including GST only as a negative control, and then subjected to Western blot probing with Pan Histone H3 antibody. The results from this show that C-PHD interacts with Histone H3, as expected, however N-PHD does not show an interaction. This suggests that this domain does not interact with any form of histone, as the source material for this experiment was the entire pool of mononucleosomes (Figure 12).

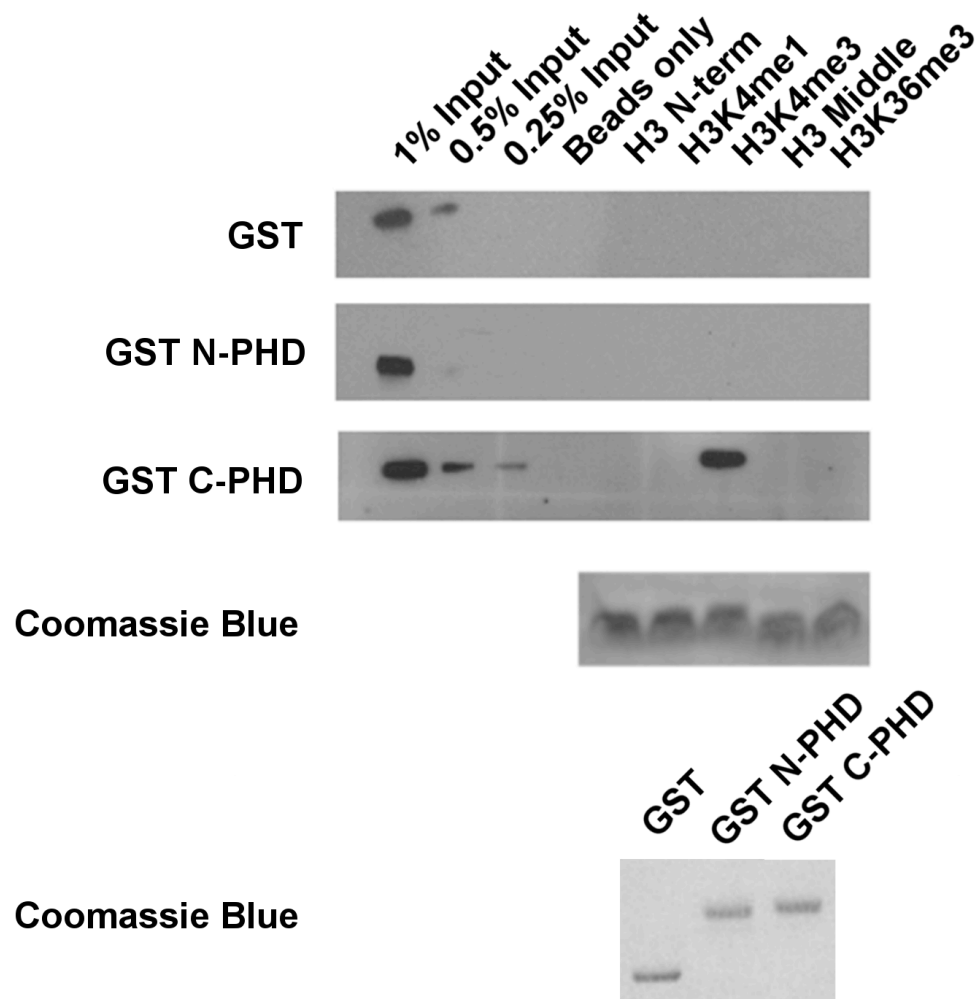


Figure 11. N-PHD Does not Recognize Select Histone Modifications. Modified histone peptide pull down indicates N-PHD does not interact with select histone modifications. The C-PHD domain recognizes H3K4me3 as previously reported (39), and serves as a positive control. The GST blot serves as a negative control in addition to the beads only pull down. Western blot probed with GST primary antibody and alpha rabbit HRP. Loading control for both the peptide sequences and purified recombinant proteins is shown.

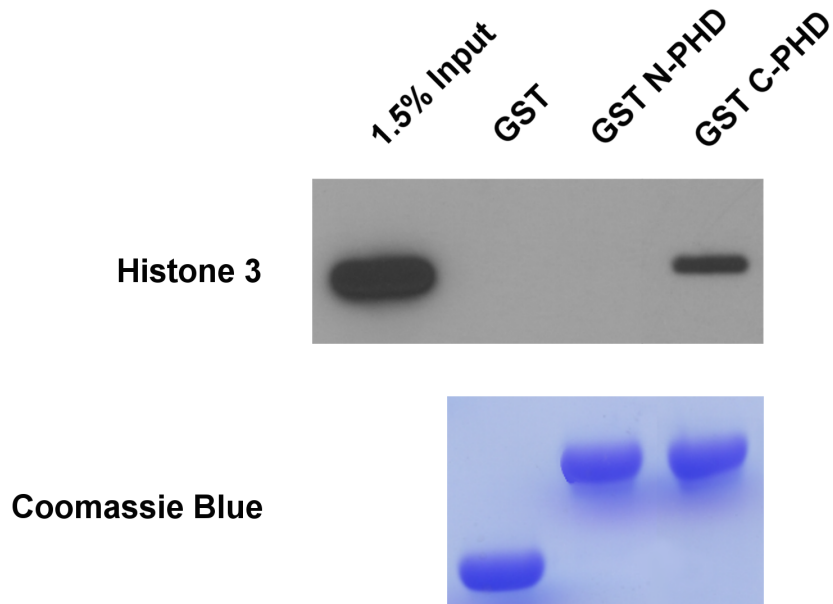
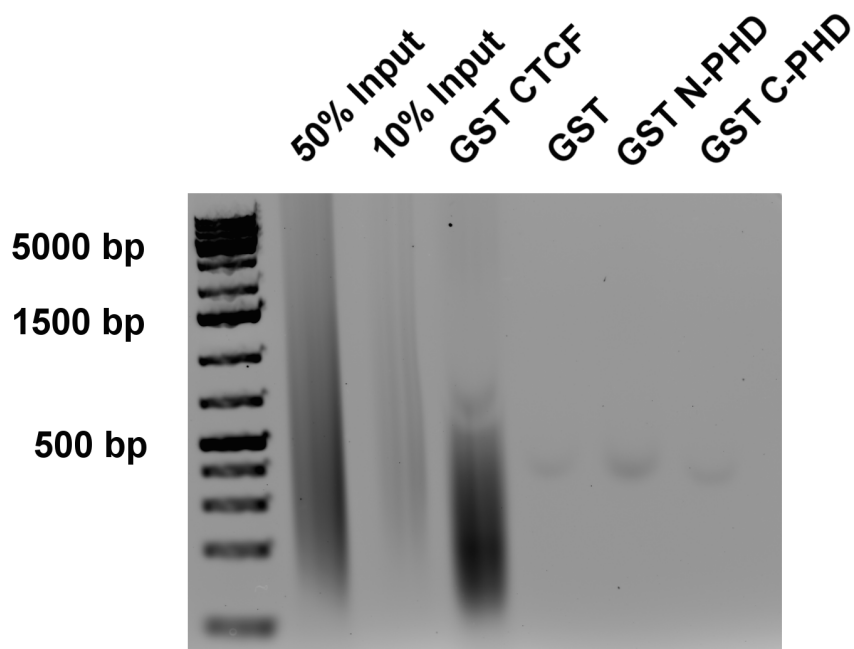


Figure 12. Bptf's N-PHD Domain Does not Interact With Histones. Western blot for MNase digest pull down experiment using GST fusion proteins for the N-terminal and C-terminal PHD finger domains of Bptf. GST alone is used as a negative control. Western blot was probed with pan Histone H3 primary antibody and rabbit HRP. The C-PHD domain shows interaction as previously reported (7). Coomassie Blue demonstrates equal loading for each of the fusion proteins.

The N-terminal PHD Finger Domain of Bptf Does not Interact With Nucleic Acids.

It was then of interest to investigate the possibility of N-PHD interacting with either DNA or RNA. As mentioned previously, PHD finger domains have been shown to interact with DNA (25). Two *in vitro* pull down experiments were performed using sonicated salmon sperm DNA and purified mRNA. For the DNA experiment, a GST fusion of CTCF was used as a positive control, and for the RNA binding experiment, MBP fusion Srsf9 was used as a positive control. CTCF is an architectural protein that is commonly found at promoters and enhancers where it functions to regulate chromatin (33). SRSF9, serine-rich splicing factor 9, is an mRNA binding protein that is involved in the splicing of mRNA transcripts. The results from both of these experiments suggest that N-PHD does not interact with nucleic acids (Figures 13 and 14).



Coomassie Blue

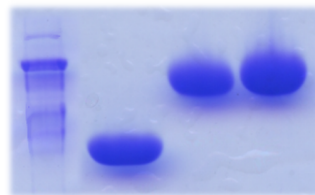
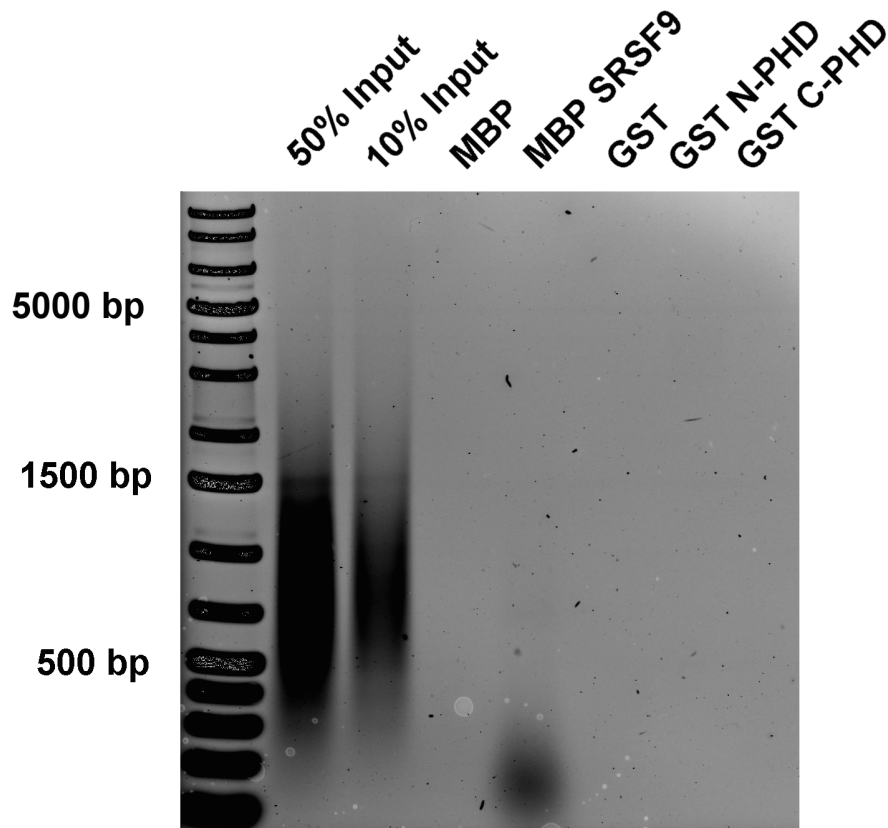


Figure 13. The N-terminal PHD Finger Domain of Bptf Does not Interact With DNA. *In vitro* DNA binding experiment using sonicated salmon sperm DNA and GST fusion proteins to investigate DNA interactions. GST fusion CTCF was used as a positive control and GST only was used for a negative control. The results show that there is no interaction between either of the PHD finger domains with DNA.



Coomassie Blue

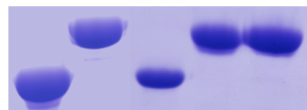


Figure 14. The N-terminal PHD Finger Domain of Bptf Does not Interact With mRNA. Results from *in vitro* mRNA pull down experiment. MBP fusion SRSF9 was used as a positive mRNA binding control. MBP and GST were used as negative controls for each of the fusion proteins, and neither of the PHD finger domains interacts with mRNA.

The N-terminal PHD Finger Domain of Bptf is Unlike Other PHD Finger Domains.

Based on the results above, it was of interest to determine the conservation of the N-PHD domain of Bptf. To do this, alignments were performed using Clustal Omega (EMBL EBI) with the N-terminal PHD finger domain of the mouse and human alongside *Danio rerio*, *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Xenopus laevis* as well as other species for a total comparison of 18 sequences (Figure 15). This analysis shows that the N-terminal PHD finger is highly conserved among vertebrates, with slight variation among the plants (Figure 15). This conservation suggests that the function of this domain may be similar across species.

It was then of interest to determine how similar the N-terminal PHD finger domain was to other PHD finger domains. To do this, the amino acid sequence for the mouse and human were run through BLASTp (NCBI) using the RefSeq protein database to determine proteins with similar PHD finger domains in both human and mouse. Through this search, an almost identical list of proteins was generated for both mouse and human (Figure 16). Further investigation of these proteins and their functions showed that the majority of them recognized histones, modified or unmodified (11). There were a few proteins in which the domain had a nonhistone protein interaction partner. Altogether, these protein domains are similar in that they are PHD finger domains and the conserved residues are those that represent the Cys₄-His-Cys₃ motif, but there is not a common ligand among them.

```

gi|571519997_Apis          LLHEGNMHYDDHCRVCHRLGDLCCETCPAVFHLECVPEPLVDVPTEDWQCSTCKAHKVTGVADCI PDVE 70
gi|281360563_Drosophila    MLQEGPIHYDDHCRVCHRLGDLCCETCPAVYHLECVDPMPNDVPTEDWQCGLCRSHKVSQVVDV---- 66
gi|694969281_Pan          LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|545857150_Sus          LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|634829668_Orycteropus  LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|530411796_Homo         LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|513213711_Gallus       LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|545178520_Equus       LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|545510109_Canis       LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|514453820_Cavia       LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|741964582_Bos         LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|594676379_Balaenoptera LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|512857301_Xenopus     LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|568972626_Mus        LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|672025368_Rattus     LMSEGVQYDDHCRVCHKLGDLLCCETCSAVYHLECVKPPLEEVPEDEWQCEVCVAHKVPGVTDCAVEIQ 70
gi|688542724_Danio      LMSEGVQYDDHCRVCHRLGDLCCETCSAVYHLECVKPPLEEVPEDEWQCEICVAHKVPGVTDCLTEFQ 70
gi|334188199_Arabidopsis  -----CVICDLGGDLCCDSCPRTYHTACLNPPLEKRIIPNGKWICPKC----- 42
gi|115472215_Oryza      -----CVECDLGGNLLCCDSCPRTYHLECLNPPLEKRAPPGNWCPCRCRTKKV----- 47
          *  *  * :*****: . : * * : * : *  *  *  *

```

Figure 15. Alignment of the N-terminal PHD Finger Domain Across 18 Species.

Alignment was done using Clustal Omega (EMBL EBI) with sequences taken from the NCBI Blastp database. The alignment shows sequence conservation among vertebrates with slight deviance in *Arabidopsis thaliana*, and *Oryza sativa*. * represents conserved amino acid residues, : represents conserved class of amino acid, and . represents a conserved amino acid residue with similar properties outside of classification.

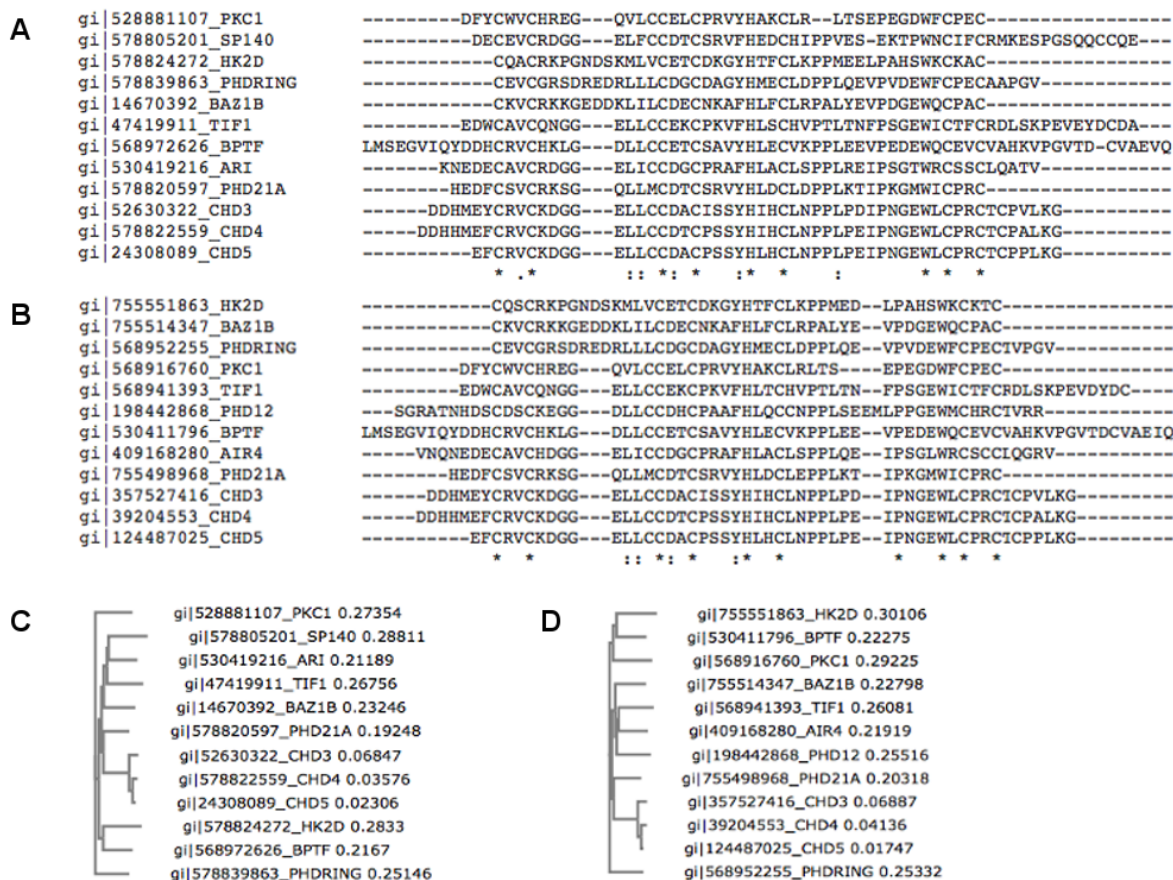


Figure 16. The N-PHD is not Similar to Other PHD Finger Domains. Alignments of the N-terminal PHD finger domain of human (A) and mouse (B) BPTF against the BLASTp databases for human and mouse, respectively. C and D show cladograms for human and mouse, respectively, indicating the relatedness of each of the proteins. * represents conserved amino acid residues, : represents conserved class of amino acid, and . represents a conserved amino acid residue with similar properties outside of classification.

The N-terminal PHD Finger Domain of Bptf Interacts with Nonhistone Proteins

The results from previous sections indicate that N-PHD of Bptf does not interact with histones or nucleic acids. Therefore, the next potential interaction partner is a non-histone protein. In order to investigate the binding partner for the N-PHD domain, an affinity column was generated with GST and GST N-PHD proteins, and nuclear extract prepared from P19 cells then ran through the column. The resulting bound proteins were eluted using three elutions and the resulting fractions were precipitated with TCA, and run on either 4-20% gradient or 15% Tris-glycine acrylamide gels. The proteins were visualized using silver staining, and the gel was then analyzed for differences between the experimental and control samples for each fraction. As seen in Figure 17, there is a difference found in experimental (GST N-PHD) and control (GST) samples between 25kDa and 37kDa and just above 75kDa in the first two washes. These bands were excised and analyzed via mass spectrometry to identify the protein (work done by Charles Lyons, Dr. Moran's lab, VCU). From this analysis it was determined that the band was primarily composed of THO complex subunit 4 (Figure 18, Table 2) and high mobility group protein B1 (Figure 19, Table 2). There were many proteins identified through this analysis, and as a starting place Hmgb1 and Thoc4 were investigated based on their biology. These two proteins were also identified in previous replicates of this experiment, providing further evidence of an interaction.

To confirm that the interaction is specific, different experiments were performed. First, a Western blot was performed on the column fractions using an antibody specific to Thoc4 and the Hmgb1 proteins. From this, it was seen that the interaction was specific to the N-PHD domain, as it was not found in the control, GST, lanes (Figure

20). The interaction was confirmed *in vivo* using nuclear extract from CJ7 cells and an antibody for Thoc4 (data not shown). The *in vivo* experiment for Hmgb1 is still in progress.

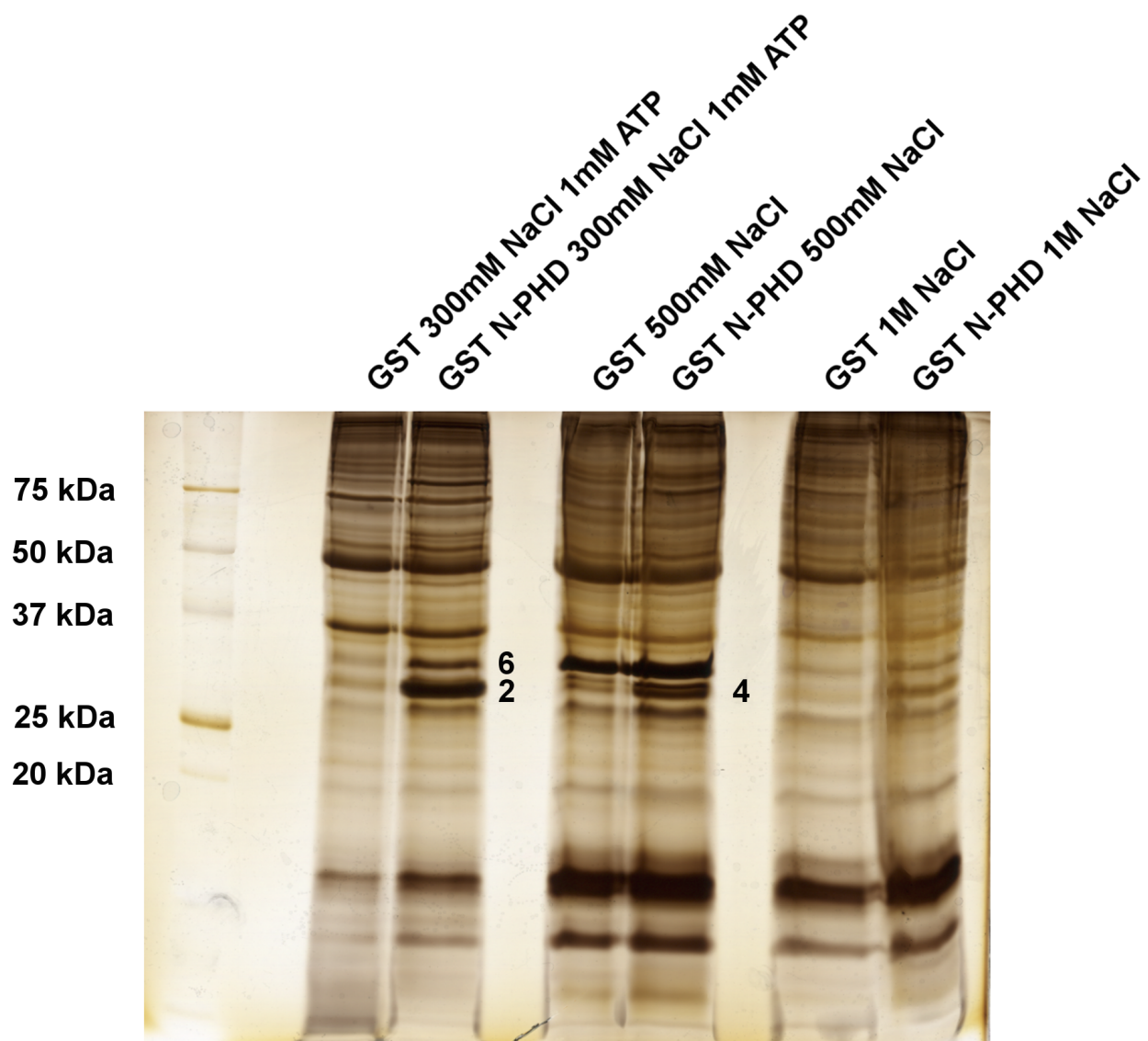


Figure 17. The N-PHD Domain Interacts With Nuclear Proteins. Silver Stained 15% Tris-Glycine acrylamide gel from pull down performed with GST and GST N-PHD linked resin with P19 nuclear extract. Numbers on gel correspond to samples presented in Table 2. The corresponding space in the GST lane was run as a control in each case for the mass spectrometry analysis.

Table 2. Proteins Identified as Potential Interaction Partners.

Sample #	# Peptides	% Coverage	Gi	Protein
2	5	29.8	6754208	High mobility group protein B1
	2	7	6755763	THO complex subunit 4
	3	21.4	6680229	High mobility group protein B2
4	10	50.2	6754208	High mobility group protein B1
	4	17.7	6755763	THO complex subunit 4
	2	14.6	658132456	Chromatin target of PRMT1 protein isoform 6
	4	14.5	25458810	Histone H1.3
	2	10.3	21426893	Histone H1.5
6	3	10.1	439253893	Actin alpha skeletal muscle
	3	15.3	6755763	THO complex subunit 4
	2	10	7242171	Proliferating cell nuclear antigen

Gi: protein accession number

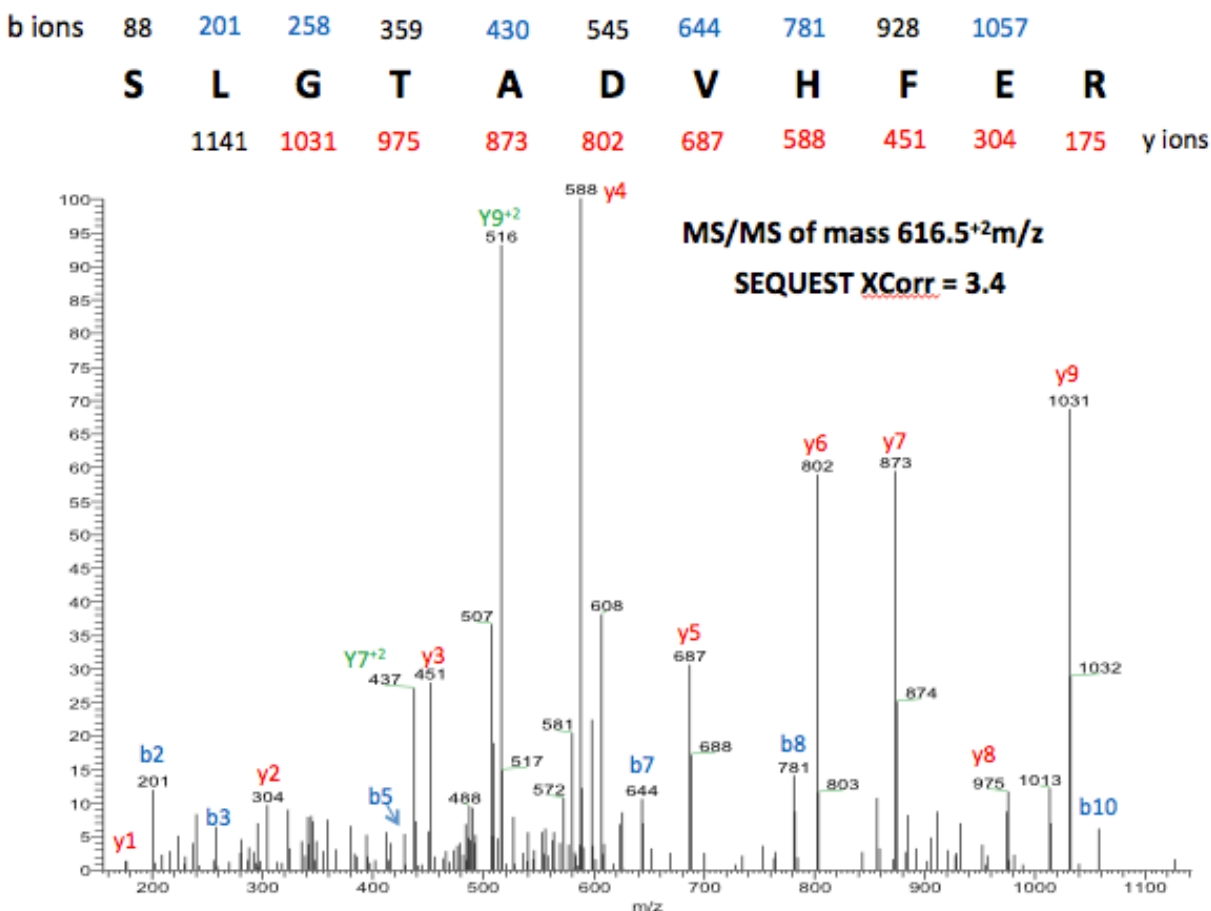


Figure 18. Representative MS² Peptide Scan for Thoc4. Scan shows ion 616.5⁺²m/z for peptide SLGTADVHFER from THO complex subunit 4. The scan shows nearly all b and y ions matched to their predicted masses. The Sequest XCorr of 3.4 indicates high confidence.

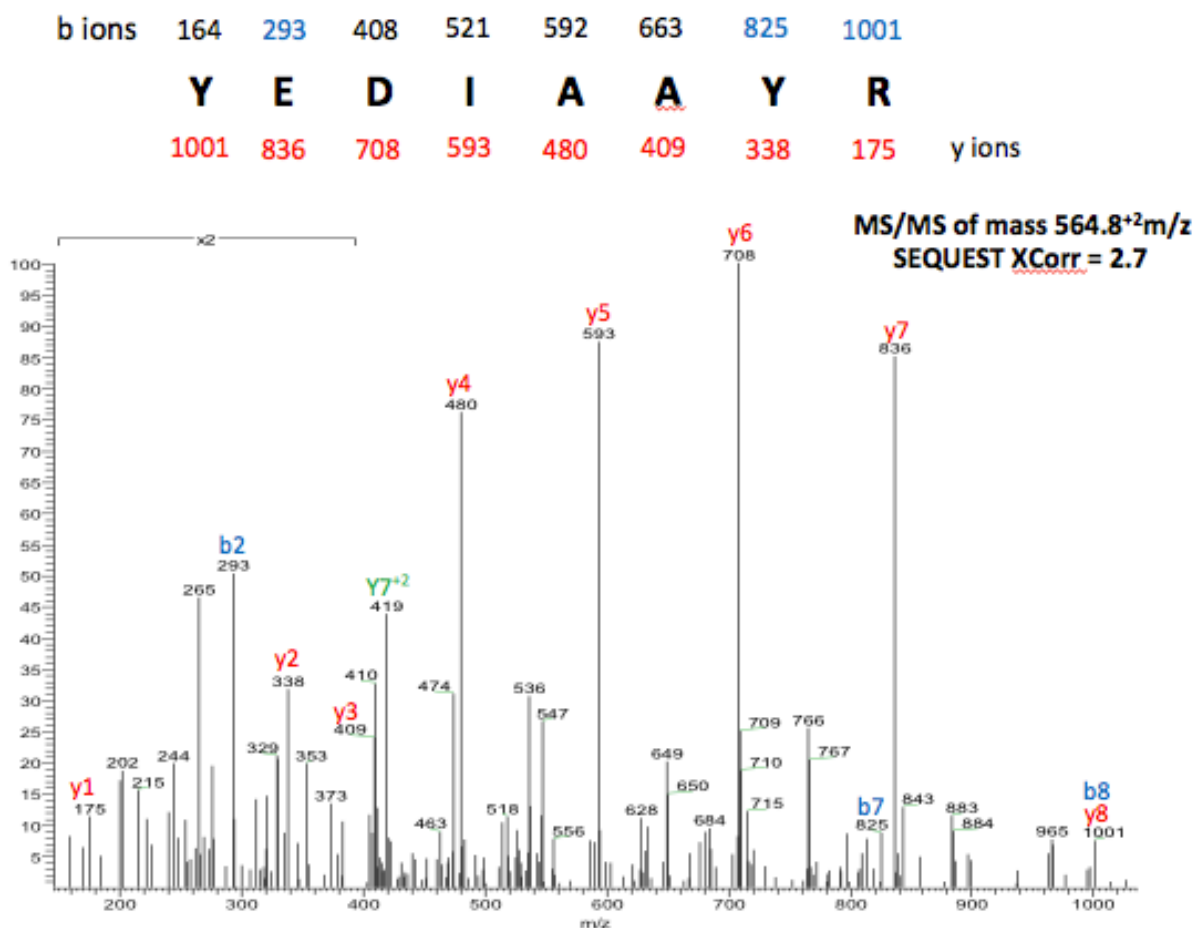


Figure 19. Representative MS² Peptide Scan for Hmgb1. Scan shows ion 564.8⁺²m/z for peptide YEDIAAAYR from Hmgb1. The scan shows some of the b ions and all of the y ions matched to their predicted masses. The Sequest XCorr of 2.7 indicates high confidence.

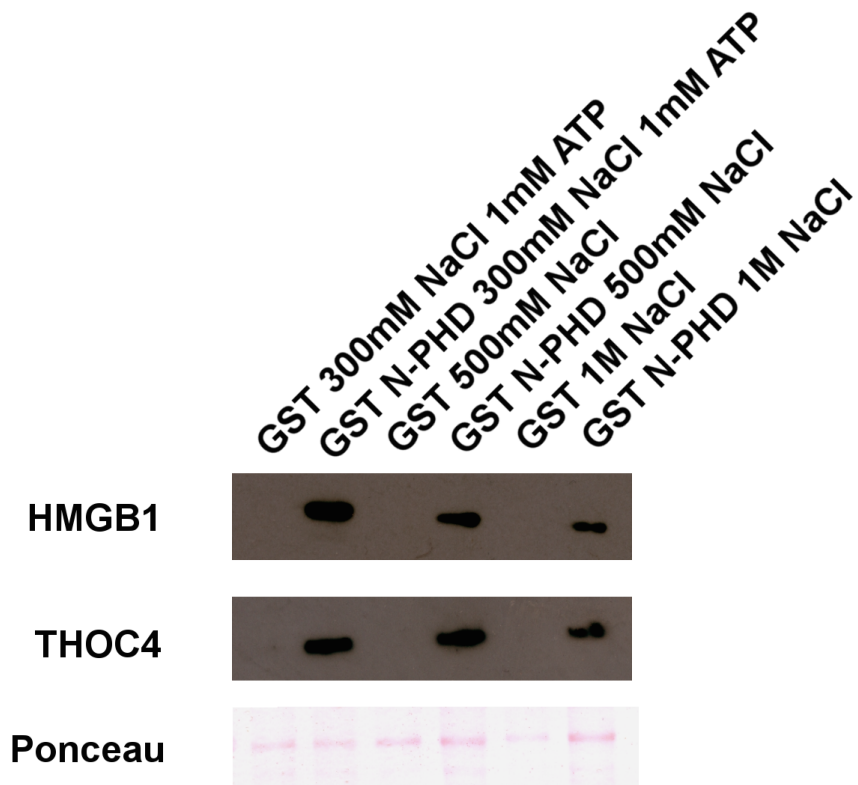


Figure 20. Hmgb1 and Thoc4 are Specific to N-PHD. Western Blot of column fractions for potential protein partners. The blot was probed with antibody against Hmgb1 and Thoc4. Ponceau staining shows equal loading among the respective column fractions.

Chapter 4: Discussion

The Bptf cDNA is Unable to be Expressed in Mouse ES Cells.

The exogenous expression of proteins is a widely used technique to study their role in various cellular processes through the use of reporters or mutant forms. This technique has been used in a variety of cell types, including ES cells. In a study done by Kaneko, et al, investigating the role of Jarid2 (member of the Polycomb repressor complex that facilitates formation of heterochromatin) on ES cell differentiation they were able to express constructs representing different mutations of Jarid2 in mouse ES cells (21). Following the expression of these constructs, the researchers were able to perform RNA immunoprecipitation experiments to determine the interaction surface of the exogenous protein with RNA (21). It has also been shown that multiple mutant forms of a protein can be expressed in ES cells and used in immunoprecipitation experiments to determine whether acetylation of specific residues in a protein were important for the recruitment of a SWI/SNF remodeling complex to the HIV promoter (26). These studies show that exogenous expression of proteins and their mutant forms is a powerful technique to study the interactions of proteins with their targets and further understand the mechanisms involved.

In attempts to solve the expression problem for our Bptf cDNA, many variables have been tested. These include the promoter, tag, different delivery methods, as well

as an analysis for protease signals in the cDNA sequence itself. For the promoter both EF1- α and CMV promoters were tested, and neither of them made a difference. The current promoter in the cDNA is the CMV promoter, and this is the same promoter that the GFP used alongside the nucleofection is expressed under. As shown in Figure 8, GFP is easily expressed in the ES cells. In addition to the promoter, the purification or identification tag on the protein was investigated. Tested tags include cMyc, 6X His, biotin, and V5. None of these tags seemed to allow the expression of this exogenous cDNA. The delivery method was also treated as a variable and both transfection and nucleofection were tested. Transfection using Lipofectamine 2000 (Life Technologies) and Nucleofection (Lonza) were tested. The transfection method involves the coupling of plasmid DNA to lipid molecules, which are then transported to the nucleus of the cell (3). Nucleofection involves electroporation, which uses electrical pulses to generate pores in the cell membrane that allows DNA to move into the cell (2). Finally, a PEST search was performed. This program does a search for known sequences recognized for proteolytic degradation (34). These sequences are rich in proline, glutamate, serine, and threonine, (PEST) are generally 12 amino acids or less in length, and tend to be hydrophilic (34). The results from this analysis did not indicate any known PEST sequences in the cDNA. All of these variables proved to be ineffective at enhancing the expression of the cDNA. The cDNA was also tested in a human cell line, 293T, and it was able to be expressed and produce exogenous Bptf (Figure 10). Therefore, the expression of the Bptf cDNA is not completely restricted in mammalian cells, however the expression in mouse ES cells appears to be inhibited in some way.

Through the testing of many variables, it has been determined that the Bptf cDNA generated in our lab is unable to be exogenously expressed in mouse ES cells. In this study, I tested the insertion of an additional sequence to our cDNA that was found in the human version of BPTF. The cDNA was transfected into CJ7 mouse ES cells, and expression was tested through a streptavidin-HRP Western blot, as the cDNA has a biotin tag. Through this, it was found that there was no expression of our exogenous Bptf. A possible explanation for this could be that the transcript is unstable in mouse ES cells. In addition to this, an issue might arise in the translation into protein, targeting these proteins for degradation either during or shortly after they are translated. There must be something that we have not detected ourselves that is targeting this transcript or protein for degradation.

The ability to express this cDNA will be important for future studies in our lab to investigate the recruitment of the NURF complex to specific sites of the genome. We ultimately aim to test the role of the different domains of Bptf in the recruitment of NURF to sites in the genome identified by the genome wide data in the lab. The next step in being able to test these mutations will be to use the CRISPR/Cas9 system to generate mutant cell lines and use these to measure NURF localization through ChIP experiments. It is important to generate these mutant Bptf proteins in order to dissect the specific recruitment of the NURF complex to specific sites found in the genome.

The N-terminal PHD finger is not a standard PHD finger domain

PHD finger domains are typically defined as recognizing histone tails in their modified or unmodified states. These domains have been partitioned into roughly seven distinct groups based on the proteins they recognize, with the possibility of more groups

to be identified. These groups include PHD finger domains that recognize H3K4me3, H3K4 (unmodified), H3K9me3, H3K36me3, H3K14Ac, non-histone proteins, and nucleic acids (29, 40, and 24). It has been suggested that in addition to the characteristic Cys₄-His-Cys₃ motif there are other residues or factors that may play a role in the interactions of PHD finger domains (24). This is not surprising as there is little sequence similarity among PHD finger domains, and with more evidence showing that these domains have diverse interactions it is likely that other important residues will be identified that further classify these protein domains.

In this study, I determined that the N-terminal PHD finger domain of Bptf does not represent a common PHD finger domain. The first potential partner I tested was histones. To test this, two experiments were performed, a pull down using GST fusion proteins and CJ7 extract generated from MNase digest, and a peptide pull down experiment using selected histone modifications and purified GST fusion proteins. Through both of these experiments, it was shown that N-PHD does not recognize histones. The C-PHD domain of Bptf was used as a positive control in both of these experiments. The MNase digestion was performed because it allows for the generation of mononucleosomes (31). MNase is a nuclease that cuts DNA at the 5' end of A/T base pairs, and when used in high enough concentration is able to digest all of the linker DNA between nucleosomes resulting in single nucleosome particles (31). The finding that the N-terminal PHD finger does not bind histones is interesting in that many proteins with similar PHD finger domains in Figure 16 are known to bind histones. The histone modifications tested were based on the genome wide data in the lab. Through this data it is seen that there is overlap of Bptf and H3K36me3, which is an indicator of

transcriptional elongation. It is also seen that there is a large amount of overlap of Bptf with H3K4me3, a marker of active promoters, and H3K4me1, which is a marker of active enhancers (20). Therefore, from this it can be said that the N-terminal PHD finger domain of Bptf does not recognize histones and that another domain must be responsible for the recruitment of Bptf to these sites.

The next binding partner tested was nucleic acids. There have been few reports of PHD finger domains interacting with nucleic acids, however it has been shown that these domains can interact with DNA, and possibly RNA (25). Two experiments were performed to test the binding of N-PHD to nucleic acids. GST and MBP fusion proteins were generated and used in an *in vitro* pull down using either sonicated salmon sperm DNA or purified mRNA. Using sonicated salmon sperm DNA should ensure that there is an equal representation of potential binding sequences added to each reaction. The purified mRNA should also contain enough representative sequence that an interaction would be detectable should it exist. The results from these experiments suggest that the N-PHD domain of Bptf does not bind to DNA or RNA. The PHD finger that has been shown to bind DNA was shown to have a large amount of basic amino acid residues (Arg and Lys), which prompted the researchers to investigate a DNA binding partner (25). The N-PHD of Bptf has a large amount of acidic residues on it (~20%), therefore these results are not surprising as negatively charged DNA would attract a positively charged partner. It should be noted that the C-PHD of Bptf has a roughly equal percentage of acidic residues (17%), and a higher percentage of basic residues (12% vs. 6%). These experiments helped eliminate the possibility of a binding partner composed of nucleic acids, and the differences in the percentage of basic amino acids

between the N-terminal and C-terminal PHD finger domains may provide further insight into their distinct functions.

The experiments performed lead to only one more potential binding partner, non-histone proteins. The most likely type of protein is a chromatin-associated protein, as the other PHD finger domain-protein interactions have been with chromatin associated proteins and PHD finger domains are found in proteins known to interact with chromatin (42). To investigate a protein partner, a column was generated with the GST fused N-PHD finger domain of Bptf and a GST only control covalently linked to resin, and nuclear extract from P19 mouse embryonic carcinoma cells was poured over the column and fractions collected. Proteins were eluted using three fractions as described in the methods. The ATP elution was performed to release heat shock proteins, and the base buffer for that fraction was the same buffer the final nuclear extract was equilibrated in. Mass spectrometry was used to analyze the protein content of the bands that were found in the experimental samples compared to the control (Figure 17). Through this analysis, it was identified that the ~30kDa protein band enriched in the N-PHD lane contained Thoc4 and Hmgb1 with high confidence (Table 2). The identification of these proteins was confirmed via a Western blot on the samples from the column experiment in Figure 16 (Figure 20). This specific interaction was confirmed via *in vivo* pull down using an antibody to Thoc4, Hmgb1 is still in progress (data not shown). The *in vitro* interaction still needs to be confirmed via a pull down using purified MBP fusion Thoc4 and Hmgb1 proteins and GST fusion proteins for wild type and mutant N-PHD to determine if the interaction is direct.

Two Potential Binding Partners for Bptf Identified

Through these experiments, it was identified that the N-terminal PHD finger domain of Bptf recognizes proteins. Through the analysis done, potential binding partners include Thoc4 and Hmgb1. Further work needs to be done in order to determine if both of these interactions are direct and to determine how these proteins play a role in the recruitment of Bptf to specific sites in the genome.

The Thoc4 protein is a nuclear protein that is a member of the TREX (transcription export) complex (36). This complex is composed of many proteins involved in transcription as well as the export of mRNA from the nucleus and includes Thoc4, Uap56, and Cip29 (22). The complex has been shown to move along actively transcribed genes with the RNA polymerase II transcription machinery (22). In yeast, this complex has been found at all actively transcribed genes, while in higher eukaryotes it is found at specific genes including those that play a role in development and differentiation (22). The TREX complex has been associated with elongation of mRNA transcripts and processing of the 3' end (22). The complex is thought to be loaded on active genes; this is believed to happen through interactions of the complex subunits with RNA polymerase II (22). The Thoc4 protein is then released to the mRNA transcript where it serves as a marker for mRNA export proteins to assemble (22). It has been suggested that the release of Thoc4 and formation of the nuclear export complex occur simultaneously (22). In addition to this, recent studies have suggested that Thoc4 has a bigger role outside of mRNA export. RNA-Seq analysis showed that Thoc4 directly affects numerous transcripts and that loss of Thoc4 results in reduced transcript stability for the selected representative transcripts (36). It was also shown that this

protein binds directly to the target transcripts, which suggests Thoc4 may directly impact their expression (36). Thoc4 is a diverse protein that is involved in transcription, processing of the 3' end of nascent mRNA, and the assembly of mRNA for nuclear export.

Based upon the genome wide data in the lab, it has been shown that NURF is found in actively transcribed genes; specifically that it may play a role in transcriptional elongation. My experiments suggest that Bptf interacts with Thoc4 through its N-terminal PHD finger domain. As mentioned above, Thoc4 is recruited to active genes through its interactions with the TREX and THO complexes and follows RNA polymerase II machinery until it is released to the mRNA transcript (22). A possible mechanism for the interaction between NURF and Thoc4 could be that NURF remodels the chromatin at active genes, allowing for transcription activation, and through its interaction with RNA polymerase II, Thoc4 interacts with Bptf to stabilize the complex of transcriptional machinery. It is also possible that Thoc4 specifically recruits NURF to certain genes, as it has been suggested that the TREX complex specifically regulates the expression of genes involved in differentiation and development in higher eukaryotes and Bptf has been shown to play a role in development (Figure 21A) (22, 18).

Hmgb1 is a non-histone nuclear protein that interacts with chromatin to alter its structure (38). This protein plays a role in transcription through its ability to bind DNA and transcription factors, and it has also been shown to interact with the RNA Pol II transcriptional machinery (38). Hmgb1 can bind to the linker DNA of the nucleosome through which it is able to “prime” the nucleosome with its ability to bend the DNA (38).

This bending of the DNA loosens the DNA-histone bond and allows for transcription factors and even chromatin remodeling complexes to come and bind to the DNA (38). Through these interactions, Hmgb1 is able to enhance transcription. Watson et al., have shown that Hmgb1 interacts with the N-terminal tail of H3 and they suggest that through this interaction, the nucleosome structure is relaxed, opening up the DNA for further interactions with transcription factors or chromatin remodeling complexes (38). Hmgb1 has been described as a non-histone chromatin associated protein and through its interactions with DNA and transcription factors it has the potential to play an important role in gene expression.

Hmgb1 functions as an architectural protein through its ability to bind to DNA and facilitate the formation of nucleoprotein structures at the nucleosome. The specific recruitment of Hmgb1 to the genome has not been characterized; it is known that this protein prefers bent DNA (12). However, a potential mechanism for its involvement with NURF can be postulated. It has been shown by Bonaldi, et al. that Hmgb1 improves the ability of ACF, an ISWI family chromatin remodeler, to bind to the nucleosome (12). Based on this information, it can be hypothesized that Hmgb1 binds to the linker DNA resulting in a loosening of the DNA wrapped around the histone. NURF could then be recruited to the site of loose DNA by Hmgb1 and through its interaction with this protein, will be able to facilitate sliding of the nucleosome and result in activation of gene expression (Figure 21B).

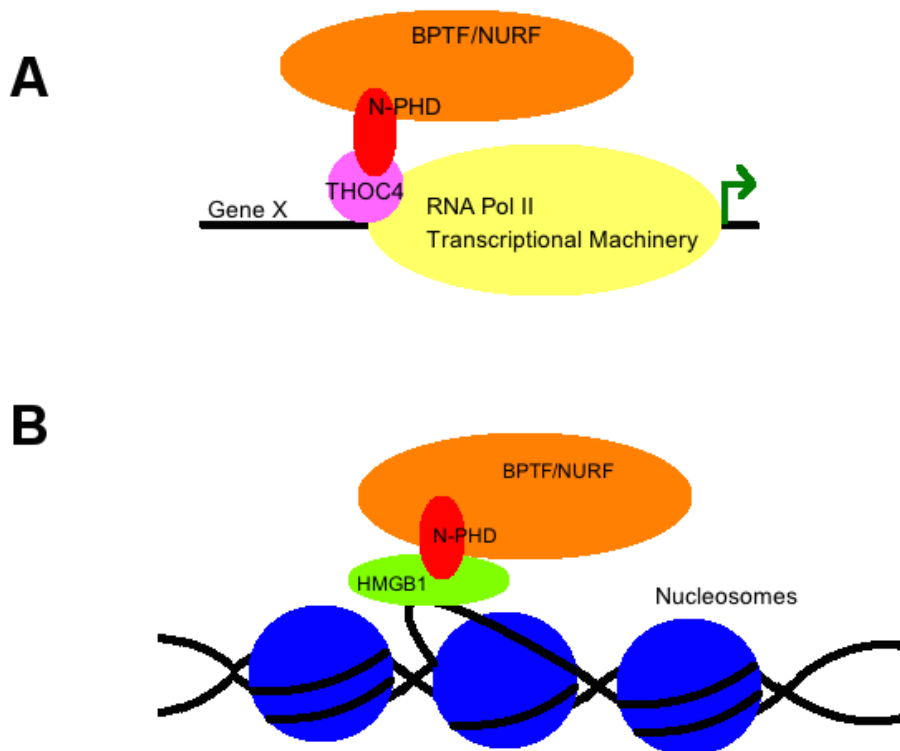


Figure 21. Model for Interaction Between NURF and Thoc4 (A) and Hmgb1 (B). Orange oval represents NURF, red oval represents N-PHD, pink oval represents Thoc4, yellow oval represents RNA Polymerase II machinery, dark blue circles represent nucleosomes, and green oval represents Hmgb1.

Conclusions

Overall, the data I have presented here suggests that the N-terminal PHD finger domain of Bptf interacts with a non-histone protein, and is not a traditional PHD finger as defined above. This finding provides further evidence that PHD finger domains are a diverse group of protein domains that interact with chromatin through either histones or non-histone protein interactions (11). The N-terminal PHD finger of Bptf has been shown to interact with two chromatin-associated proteins, and their distinct functions

may provide further insight to the recruitment of the NURF complex to different regions of the genome.

Further investigation needs to be done to completely understand the interaction of Bptf with either Thoc4 or Hmgb1. The interactions need to be confirmed via *in vitro* pull down, this will provide evidence that the interaction is direct. It will also need to be confirmed that this interaction can be abolished with a specific mutation, such as one of the two mutant N-PHD proteins available in the lab. Genome wide data of Thoc4 or Hmgb1 recruitment in mouse ES cells will provide information on the overlap with NURF in the genome and further describe the interaction between these two proteins. If specific sites are identified to overlap between the two proteins, performing ChIP in knockdown ES cells will provide information about how these proteins specifically impact the transcription of genes. In addition to this, the interaction surface between the N-terminal PHD finger domain and Thoc4 or Hmgb1 can be further dissected and the molecular mechanism determined. It may also be of interest to pursue the other proteins identified through the mass spectrometry analysis. Investigation of these proteins will further elucidate how NURF is recruited to specific sites in the genome.

I have also investigated the expression of a cDNA for Bptf, and found that it is unable to be expressed in mouse ES cells. It will be important to generate mutant Bptf proteins that can be exogenously expressed in mouse ES cells to not only confirm the results from our genome wide data, but to also further dissect the recruitment of the NURF complex to specific sites identified in the genome. It is not only important to identify genes or regions in the genome that specifically recruit Bptf/NURF, but it is

important to fully understand the interaction surface and the particular domains involved.

Ultimately, understanding the recruitment of the NURF complex to the genome will allow the lab to determine whether or not to pursue the complex as a therapeutic target. It will be important to fully understand its recruitment to determine what types of genes will be affected by a Bptf/NURF inhibitor or how elevated levels of this complex will impact a healthy versus abnormal cell.

Literature Cited

1. "ATP-Dependent Chromatin Remodeling Complexes as Novel Targets for Cancer Therapy." *Advances in Cancer Research*. Ed. Kenneth Tew and Paul Fisher. Vol. 121. Elsevier Science, 2010. 183-233.
2. "Improved Electroporation Using Nucleofector Technology." *Lonza*. Web. 8 Apr. 2015. <<http://www.lonza.com/products-services/bio-research/transfection/nucleofector-technology.aspx>>.
3. "Plasmid DNA Transfection." *Life Technologies*. Web. 8 Apr. 2015. <<http://www.lifetechnologies.com/us/en/home/life-science/cell-culture/transfection/choosing-a-transfection-reagent.html>>.
4. Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. Chromosomal DNA and Its Packaging in the Chromatin Fiber. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK26834/>
5. Alberts, Bruce. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science, 2002.
6. Alkhatib, Suehyb G., and Joseph W. Landry. "The Nucleosome Remodeling Factor." *FEBS Letters* 585 (2011): 3197-207.

7. Badenhorst, Paul, Matthew Voas, Ilaria Rebay, and Carl Wu. "Biological Functions of the ISWI Chromatin Remodeling Complex NURF." *Genes & Development* (2002): 3186-198.
8. Baker, Lindsey A., C. David Allis, and Gang G. Wang. "PHD Fingers in Human Diseases: Disorders Arising from Misinterpreting Epigenetic Marks." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 647 (2008): 3-12.
9. Bartke, T., J. Borgel, and P. A. Dimaggio. "Proteomics in Epigenetics: New Perspectives for Cancer Research." *Briefings in Functional Genomics* 12.3 (2013): 205-18.
10. Biegel, Jaclyn A., Tracy M. Busse, and Bernard E. Weissman. "SWI/SNF Chromatin Remodeling Complexes and Cancer." *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (2014): 350-66.
11. Bienz, Mariann. "The PHD Finger, a Nuclear Protein-interaction Domain." *Trends in Biochemical Sciences* 31.1 (2006): 35-40.
12. Bonaldi, Tiziana, Gernot Langst, Ralf Strohner, Peter Becker, and Marco Bianchi. "The DNA Chaperone Hmgb1 Facilitates ACF/CHRAC-dependent Nucleosome Sliding." *The EMBO Journal* 21.24 (2002): 6865-873.
13. Bridger, Haley. "Epigenomics Approach Illuminates the Dark Corners of the Genome." *Broad Institute*. 5 Aug. 2010. Web. 8 Apr. 2015.
<<https://www.broadinstitute.org/news/1504>>.

14. Buganim, Yosef, Ido Goldstein, et al. "A Novel Translocation Breakpoint within the BPTF Gene Is Associated with a Pre-Malignant Phenotype." *PLoS ONE* 5.3 (2010).
15. Cherry, Christopher M., and Erika L. Matunis. "Epigenetic Regulation of Stem Cell Maintenance in the Drosophila Testis via the Nucleosome-Remodeling Factor NURF." *Cell Stem Cell* 6 (2010): 557-67.
16. Dar, Altaf, et al. "The Role of BPTF in Melanoma Progression and in Response to BRAF-Targeted Therapy." *Journal of the National Cancer Institute* 107.5 (2015).
17. Dignam, John David, Lebovitz, R.M., Roeder, R.G. "Accurate Transcription by RNA Polymerase II in a Soluble Extract from Isolated Mammalian Nuclei." *Nucleic Acids Research* 11.5 (1983): 1475-489.
18. Erdel, Fabian, and Karsten Rippe. "Chromatin Remodeling in Mammalian Cells by ISWI-type Complexes - Where, When and Why?" *The FEBS Journal* 278 (2011): 3608-618.
19. Goller, T., F. Vauti, S. Ramasamy, and H.-H. Arnold. "Transcriptional Regulator BPTF/FAC1 Is Essential for Trophoblast Differentiation during Early Mouse Development." *Molecular and Cellular Biology* 28.122 (2008): 6819-827.
20. Hon, G. C., R. D. Hawkins, and B. Ren. "Predictive Chromatin Signatures in the Mammalian Genome." *Human Molecular Genetics* 18.2 (2009): R195-201.
21. Kaneko, Syuzo, Roberto Bonasio, Ricardo Saldaña-Meyer, Takahaki Yoshida, Jinsook Son, Koichiro Nishino, Akihiro Umezawa, and Danny Reinberg.

- "Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin." *Molecular Cell* 53 (2014): 290-300.
22. Katahira, Jun. "MRNA Export and the TREX Complex." *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* (2012): 507-13.
23. Landry, Joseph, Alexei A. Sharov, Yulan Piao, Lioudmila V. Sharova, Hua Xiao, Eileen Southon, Jennifer Matta, Lino Tessarollo, Ying E. Zhang, Minoru S. H. Ko, Michael R. Kuehn, Terry P. Yamaguchi, Carl Wu, and Jeannie T. Lee. "Essential Role of Chromatin Remodeling Protein Bptf in Early Mouse Embryos and Embryonic Stem Cells." *PLoS Genetics* 4.10 (2008).
24. Li, Y., and H. Li. "Many Keys to Push: Diversifying the 'readership' of Plant Homeodomain Fingers." *Acta Biochimica Et Biophysica Sinica* 44 (2011): 28-39.
25. Liu, Lei, Su Qin, Jiahai Zhang, Peng Ji, Yunyu Shi, and Jihui Wu. "Solution Structure of an Atypical PHD Finger in BRPF2 and Its Interaction with DNA." *Journal of Structural Biology* 180 (2012): 165-73.
26. Mahmoudi, T, et al.. "The SWI/SNF Chromatin-remodeling Complex Is a Cofactor for Tat Transactivation of the HIV Promoter." *Journal of Biological Chemistry* 281.29 (2006): 19960-9968.
27. Mann, Jeffrey. "Epigenetic's and Memigenetics." *Cellular and Molecular Life Sciences* 71 (2014): 1117-122.
28. Morrison, Ashby, and Xuetong Shen. "Chromatin Remodeling Beyond Transcription: The INO80 and SWR1 Complexes." *Molecular Cell Biology* 10 (2009).

29. Musselman, C. A., and T. G. Kutateladze. "Handpicking Epigenetic Marks with PHD Fingers." *Nucleic Acids Research* 39.21 (2011): 9061-071.
30. Musselman, C. A., and T. G. Kutateladze. "PHD Fingers: Epigenetic Effectors and Potential Drug Targets." *Molecular Interventions* 9.6 (2009): 314-23.
31. Nikitina, Tatiana, Difei Wang, Misha Gomberg, Sergei A. Grigoryev, and Victor B. Zhurkin. "Combined Micrococcal Nuclease and Exonuclease III Digestion Reveals Precise Positions of the Nucleosome Core/Linker Junctions: Implications for High-Resolution Nucleosome Mapping." *Journal of Molecular Biology* (2013): 1946-960.
32. Petty, Emily, and Lorraine Pillus. "Balancing Chromatin Remodeling and Histone Modifications in Transcription." *Trends in Genetics* 29.11 (2013): 621-29.
33. Qiu, Zhijun, Carolyn Song, Navid Malakouti, Daniel Murray, Ayman Hariz, Mark Zimmerman, Derek Gyax, Aiman Alhazmi, and Joseph Landry. "Functional Interactions Between NURF and CTCF Regulate Gene Expression." *Molecular and Cellular Biology* 35.1 (2015): 224-37.
34. Rechsteiner, Martin, and Scott W. Rogers. "PEST Sequences and Regulation by Proteolysis." *Trends in Biochemical Sciences* (1996): 267-71.
35. Ruthenburg, Alexander J., Haitao Li, Thomas A. Milne, Scott Dewell, Robert K. McGinty, Melanie Yuen, Beatrix Ueberheide, Yali Dou, Tom W. Muir, Dinshaw J. Patel, and C. David Allis. "Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions." *Cell* 145 (2011): 692-706.

36. Stubbs, Sarah H., and Nicholas K. Conrad. "Depletion of REF/Aly Alters Gene Expression and Reduces RNA Polymerase II Occupancy." *Nucleic Acids Research* 43.1 (2015): 504-19.
37. Swygert, Sarah G., and Craig L. Peterson. "Chromatin Dynamics: Interplay between Remodeling Enzymes and Histone Modifications." *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* (2014).
38. Watson, M., K. Stott, H. Fischl, L. Cato, and J. O. Thomas. "Characterization of the Interaction between Hmgb1 and H3--a Possible Means of Positioning Hmgb1 in Chromatin." *Nucleic Acids Research* 42.2 (2013): 848-59.
39. Wysocka, Joanna, Tomek Swigut, Hua Xiao, Thomas A. Milne, So Yeon Kwon, Joe Landry, Monika Kauer, Alan J. Tackett, Brian T. Chait, Paul Badenhorst, Carl Wu, and C. David Allis. "A PHD Finger of NURF Couples Histone H3 Lysine 4 Trimethylation with Chromatin Remodeling." *Nature* 442 (2006).
40. Xiao, Hua, Raphael Sandaltzopoulos, Hih-Min Wang, Ali Hamiche, Ryan Ranallo, Kyu-Min Lee, Dragony Fu, and Carl Wu. "Dual Functions of Largest NURF Subunit NURF301 in Nucleosome Sliding and Transcription Factor Interactions." *Molecular Cell* 3 (2001): 531-43.
41. Xiao, Shuai, Longfei Liu, Xianzhou Lu, Jianwu Long, Xiaojun Zhou, and Min Fang. "The Prognostic Significance of Bromodomain PHD-finger Transcription Factor in Colorectal Carcinoma and Association with Vimentin and E-cadherin." *Journal of Cancer Research and Clinical Oncology* (2015).
42. Zucchelli, Chiara, Simone Tamburri, Giacomo Quilici, Eleonora Palagano, Andrea Berardi, Mario Saare, Pärt Peterson, Angela Bachi, and Giovanna

Musco. "Structure of Human Sp140 PHD Finger: An Atypical Fold Interacting with Pin1." *FEBS Journal* 281 (2013): 216-31.

Vita

Marissa Irene Mack was born on May 31, 1989 in Sayre, Pennsylvania, and is an American citizen. She graduated from Waverly High School, Waverly, New York in 2007. She received her Bachelor of Science in Cell and Molecular Biology from Binghamton University, Binghamton, New York in 2011. She entered the Human and Molecular Genetics program at Virginia Commonwealth University, Richmond, Virginia, where she received her Master of Science in Human Genetics in 2015.